

**Application of PCA and Hough Transform to
classify features in optical images**

Prajuab Inrawong, MEng.

**Thesis submitted to the University of Nottingham for
the degree of Doctor of Philosophy**

April 2012



**The University of
Nottingham**

Abstract

Viewing fine features of object with optical instruments have become increasingly difficult as the dimensions of many features of interest have become smaller than the traditional optical resolution limit. Examples of these features can be found in semiconductor components and biological tissues. This has caused a move to non-optical methods such as scanning electron and atomic force microscopy techniques, or optical methods combined with signal processing techniques to provide clearer images of samples. This thesis presents a method to increase the resolution of an optical system. This is achieved by using principal component analysis (PCA). Once the PCA measured the object image parameters, the new clearer image can be reconstructed based on these parameters. This process works extremely well. Various aspects of samples measured by the PCA have been investigated, such as the shift of sample, the sample with different sizes, the orientation of sample and the impact of noise. These studies show that the technique is extremely robust, and has huge potential for general usage. The thesis also contains the detail of the Hough Transform technique which was used to provide the initial parameters to the PCA. From the analysis of the technique, it is concluded that the accurate measurement of the technique can be achieved by providing adequate templates of the object image for the system.

Acknowledgements

I would firstly like to thank my supervisor, Dr CW See and Professor MG Somekh for their patient guidance throughout this course. It has been a rewarding experience taking on this PhD project, and I am grateful to have been given the opportunity to work alongside them.

I would like to thank all friends and colleagues from the iBIOS for their generous helps and supports.

I would like to thank my family for all of their support over the past 4 years.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction..... | 1 |
| 1.1 | Introduction..... | 1 |
| 1.2 | Imaging techniques and recent developments..... | 2 |
| 1.3 | Layout of thesis..... | 12 |
| | | |
| 2 | Literature review..... | 14 |
| 2.1 | Resolution Limit..... | 15 |
| 2.2 | Spectral extension and information theory..... | 19 |
| 2.2.1 | Analytic continuation and uniqueness theory..... | 20 |
| 2.2.2 | Information theory..... | 23 |
| 2.2.3 | Analytic continuation by Taylor expansion..... | 25 |
| 2.3 | Applications of spectrum extension theory..... | 26 |
| 2.3.1 | Auto-Regressive Models..... | 26 |
| 2.3.2 | Sampling theorem infrequency domains..... | 29 |
| 2.4 | Artificial Neural Networks..... | 31 |
| 2.4.1 | The ANN model..... | 32 |
| 2.4.2 | Activation Functions..... | 34 |
| 2.4.3 | Network topographies and applications..... | 35 |
| 2.4.4 | Training network..... | 38 |
| 2.4.5 | Application to optical measurement..... | 39 |
| 2.5 | Commercial microscope..... | 40 |

| | | |
|----------|---|-----------|
| 3 | Hough Transform and PCA | 43 |
| 3.1 | Hough Transform | 44 |
| 3.2 | Representation of lines into Parameter Space | 46 |
| 3.3 | Classical Hough Transform algorithm for line detection | 51 |
| 3.4 | Principal Component Analysis | 54 |
| 3.5 | Mathematics of PCA | 58 |
| 3.5.1 | Covariance Matrix | 58 |
| 3.5.2 | Eigenvectors and Eigenvalues | 60 |
| 3.6 | Computing PCA | 63 |
| 3.7 | An example of PCA | 65 |
| | | |
| 4 | Application of Hough Transform and PCA | 72 |
| 4.1 | Lines detection in grey scale image using Hough Transform | 72 |
| 4.2 | Transformation to Hough Space | 74 |
| 4.2.1 | Accumulator 1: ACC1 | 78 |
| 4.2.2 | Accumulator 1: ACC2 | 81 |
| 4.3 | Location of line in accumulator | 85 |
| 4.4 | Length and width of lines | 88 |
| 4.5 | Adjacent lines Detection | 90 |
| 4.5.1 | System noise sources | 98 |
| 4.6 | Principal Component Analysis | 101 |
| 4.6.1 | Measuring the similarity using PCA | 101 |
| 4.6.2 | Add noise to system | 119 |

| | | |
|----------|---|------------|
| 5 | Two-Dimension PCA..... | 127 |
| 5.1 | Convert line image to vector..... | 127 |
| 5.2 | Measuring line segment images..... | 130 |
| 5.3 | Non-straight line images..... | 162 |
| 5.4 | Coarse measurement using Hough Transform..... | 164 |
| | | |
| 6 | Experimental Results..... | 171 |
| 6.1 | Sample..... | 171 |
| 6.2 | Hough Transform..... | 172 |
| 6.3 | Principal Component Analysis..... | 178 |
| | | |
| 7 | Conclusion and Future Work..... | 183 |
| 7.1 | Conclusion..... | 183 |
| 7.2 | Suggestion for future work..... | 189 |
| 7.2.1 | Multi-Dimension PCA..... | 189 |
| 7.2.2 | 3D classification..... | 190 |
| 7.2.3 | Our vision..... | 191 |
| | | |
| | Appendix A..... | 193 |
| | | |
| | References..... | 196 |

Chapter 1

1 Introduction

1.1 Introduction

The simple optical microscope was first developed in 1674 [1] when the Dutchmen Hans and Zacharias Jansenn used a single lens to look at tree trunks and insects. This is considered to be the pioneer of microscopy. Finally people could see minute features or small objects that were beyond the ability of naked human eyes. There followed a period of improvement in lens design and making. This led to the reduction of aberration associated with images obtained from the earlier systems and in turn significant improvement in the image quality [2]. In 1830 J. Jackson discovered that the combination of several weaker lenses instead of using a single strong lens reduced chromatic aberration and provided good magnification without blurring of the images and so the compound microscope was born [3][4]. By 1872 Ernst Abbe [5] formulated his 'Abbe sine condition' providing the mathematical basis to describe the maximum resolution of a particular microscope. In 1897 Lord Rayleigh [6] proposed a criterion for image resolution of optical system based on the separation of diffraction limited images of point sources. The first high power objective was made by H Powell around 1896 [7]. It corrected for three wavelengths and used oil immersion in the object space to obtain a lens with numerical aperture of approximately 1.50, thus enabling objects smaller than a micron to be observed.

The ability to resolve fine features is fundamental to microscopes. As technology progresses the demand to resolve ultra fine features has also increased such as in medical, biological and nanotechnology research. However, optical microscopes have a finite resolution due to the nature of light. Limitations of standard optical microscope rely upon diffraction and the resolution cannot exceed half of an optical wavelength. Features that are smaller than half of a wavelength will result in non-propagating evanescent wave which decay exponentially and cannot be detected in the far field. Information associated with those features is therefore lost and will not be represented in the image.

1.2 Imaging techniques and recent developments

Over the years, there have been many microscopy techniques developed for different applications, some may even have resolution finer than the conventional optical microscope. They will be considered described briefly here.

- Bright Field: Bright field microscopy is the simplest of all the microscopy techniques. The sample is illuminated by white light, either from one side of the sample and observed at the other side (transmission mode), or illuminated and observed from the same side (reflection mode). Limitations of the technique include low image contrast. Also, for semi-transparent samples, the image is often masked by out of focus features. However bright field operation requires minimal sample preparation and along with the simplicity of the system setup these are the significant advantages.

- Dark field: The contrast of an image can be improved dramatically by using dark field illumination. The technique uses a carefully aligned arrangement to effect an oblique illumination of light at the sample. The imaging optics is in turn designed to minimise the quantity of directly transmitted or reflected light entering the image plane. Therefore only light scattered by the sample will be collected and contribute to the image, with the background light heavily suppressed, thus the term dark field or dark background. Dark field microscopy is a very simple yet effective technique and well suited for uses involving live and unstained biological samples, such as a smear from a tissue culture or individual water-borne single-celled organisms [8]. The quality of images obtained from this technique is very impressive. However the technique does suffer from low light intensity in the final image and is also affected by features that locate outside of the focus.
- Phase contrast: Phase contrast microscopy is a widely used technique that shows differences in refractive index of the sample as difference in contrast of the intensity image. It was invented by Frits Zernike [9][10]. It relies on changing the relative phase shift between the diffracted and the directly transmitted light, allowing them to interfere at the image plane. The contrast of the resulting image can be up to 100x better than that achieved from using bright field operation. Phase contrast is used extensively in the study of biological samples. One disadvantage of the technique is the formation of halo around the features, which can obscure finer details.

- Nomarski microscope: Another imaging technique that is sensitive to the phase structure of the sample is the differential interference contrast microscope invented by Georges Nomarski [11]. The system consists of a special prism known as the Nomarski prism, which is a modified version of the Wollaston prism. In the reflection mode, the prism splits the incoming beam into two, and also recombines them to form an interference image. The system is sensitive primarily to changes in the phase structures of the sample, hence the term differential. The contrast of the image produced by the microscope is very good and it is mainly used to study solid samples.
- Confocal microscope: The technique uses a scanning point of light instead of full sample illumination. The reflected light is focused and detected using a point detector. The point source and point detector combine to give a slightly higher lateral resolution, but a significant improvement in sectioning of the sample. The main advantage of the confocal microscope is its ability to reduce the depth of field, thus allowing serial optical sectioning of thick specimens and leading to 3D imaging of the sample. Another advantage is that the point detector would reduce or eliminate stray light from the system which would improve the image contrast. Although the confocal microscope is used mainly in biological study, it is also used in the semiconductor industry.

The above microscopy techniques are designed mainly for improving the sensitivity of the systems to minute changes in the sample structure, and to produce high contrast images. Their lateral resolutions, apart from the confocal microscope, do not

exceed that of a conventional bright field microscope. The systems introduced below, not confining to optical, all have resolution beyond the conventional microscope. They either rely on coding the sample information in a special way so that high frequency information can arrive at the image plane, or make use of very short wavelength irradiation. For many of these systems, the ultimate limit in resolution is governed by the signal to noise ratio of the system. Some of them will be discussed here.

- Structured illumination microscope (SIM) [12]: this technique was developed roughly ten years ago. In the simplest configuration, the system employs a sinusoidal intensity pattern to illuminate the sample. The zero illumination order acts similarly to that in a conventional microscope. However, the higher illumination orders would down convert the high spatial frequency components of the object into the aperture of the microscope. There is, therefore, an increased range of spatial frequency arriving at the image plane. These components are detected, separated and rearranged to produce a high resolution image. Systems of this kind can produce images of resolution two times better than the conventional ones. Recently, a new configuration has been proposed, making use of a proximity grating. In theory this system can have up to a four-fold resolution improvement over the conventional microscope. The technique is relatively new and is used mainly in the study of cellular nanostructures stained with a fluorescent marker.
- PALM and STORM: Eric Betzig [13][14] developed the photo-activated localization microscopy (PALM) [15] and Xiaowei Zhang [16] used a similar

technique to form what is called the stochastic optical reconstruction microscopy (STROM) [17][18]. Both techniques fill the sample with many fluorophores that can be photoactivated into a fluorescing state by flash of light. The light intensity is set very low so that, for each illumination, only a few fluorophores will be activated. The image captured will therefore consist of bright spots randomly dotted across the image. The locations and brightness of these spots are determined and stored as sample information for this particular exposure. This process is repeated many times until the photoactivated fluorophores cover the entire area of interest. The stored information is then used to build up an image. Since the locations of the bright spots can be determined with much higher precision than the lateral resolution of the optical system, the image thus formed would have a much better resolution. One major drawback of the technique is that, because of the many exposures required, it takes on the order of hours to collect the data.

- Wide-field structured-illumination(SI) [19] [20]. The main concept of SI is to illuminate a sample with patterned light and increase the resolution by measuring the fringes in the Moire' pattern. SI enhances spatial resolution by collecting information from frequency space outside the observation region. With several frames with the illumination shifted by some phase, it is possible to computationally separate and reconstruct the Fourier Transform (FT) image which has much more resolution information. The inverse FT returns the reconstructed image. This technique enhances the resolution by the factor of 2.

Two alternative, non-optical techniques that have extremely high imaging resolution will be described below, and they are:

- Electron microscope: electron beam with a very short wavelength is used in electron microscopy. The scanning electron microscope (SEM) focuses a high-energy beam of electrons onto the sample surface, and an image is formed by scanning the e-beam in a raster scan pattern. Using high acceleration voltage, the resolution of electron microscope can reach sub-nanometer levels. Electron microscopes are used in biological sciences and semiconductor industry, among others. Although the techniques yield excellent resolution, they have a number of drawbacks. For example, the systems are expensive to buy compare to optical microscope, and require high maintenance. The high electron beam power can cause damage to the sample unless great care is taken. Another major disadvantage is that the microscope operates only in vacuum. The specimens also require extensive preparation [21] for them to be suitable for inspection.
- Scanning probe microscope (SPM)[61][62]: Scanning probe microscopy is another high resolution non-optical technique. There are many types of SPM and they all employ a fine solid probe in the vicinity (near field) of the object. The system output is governed by the interaction between the probe and the sample, and scanning the sample, an image if formed. Examples of the SPM include the atomic force microscope (AFM) [63], the scanning tunnelling microscope (STM) and the photonic force microscope. The atomic force microscope relies on the changes in the atomic force existed between the tip

of the probe and the sample surface. When scanned, the AFM provides a true three-dimensional surface profile, with a lateral resolution in the nanometer scale. Moreover, samples examined by AFM do not require special treatments such as metal or carbon coating. Most AFM modes can work perfectly well in air or even liquid environment. This makes it possible to study biology and even living organisms. The selection of the scanning tip is important, as an inappropriate choice may lead to image artefacts. AFM is also relatively slow, requiring typically several minutes to form an image. This may lead to thermal drift during the image formation making AFM less suited for measuring accurate distances between topographical features in an image.

The typical values for the resolution of the above microscopy methods are summarised in Table 1.

Table 1 Microscopy Methods and resolution

| Method | Resolution (nm) |
|--|--|
| Dark field, Bright field, Phase contrast Differential interference contrast microscope | ~ 230 (lateral) [67] ~ 10 (vertical) |
| Confocal microscope | ~ 180 (lateral) [67] ~ 500 (vertical) |
| Structured light microscope (SIM) | ~ 120 (lateral) [67] |
| Photo-activated localization microscopy (PALM) Stochastic optical reconstruction microscopy (STORM) | < 20 (lateral) [67] |
| Scanning electron microscope (SEM) | ~ 10 (lateral) [68] |
| Scanning probe microscope (SPM) | ~ 10 (lateral) [69] ~ 5 (vertical) |

As it is becoming more demanding to keep up with the technological advances in industry and research, a method to efficiently measure small features of object below the optical resolution limitation is necessary. There have been numerous approaches [22][23][24][25] to address this problem. Many of them tackle the problem by returning to the underlying mathematics that is used to model the relationship between the object and the image formation mechanism. They attempt to reverse the effect of the optical system imposed onto the sample and retrieve the information lost in the imaging process. An example is based on the application principle of analytic continuation, and that the spectrum of a bounded object is an analytic function [22]. By applying appropriate algorithm, the spectral components originally outside the bandwidth of the microscope are recovered, thus leading to an image that has higher resolution than that can be obtained directly from the microscope. This allows the construction of a super-resolved image and enables much smaller object features to be observed than with a conventional microscope. This process works to some extent for samples that are simple and well behaved, and that the optical system is near perfect and does not suffer from aberrations. Furthermore, the signal generated should be noiseless. When dealing with a real system, however, none of these conditions can be satisfied, and a small amount of noise would render the approach totally ineffectual.

Recently there has been research in this area which uses a very different approach. Instead of obtaining images of very high resolution, they aim to measure the dimensions of some selected features accurately. The information thus obtained is used to make up a high resolution image. The rationale behind such an approach is based on the theory of information content, that the total information associated with

an image is fixed. Any attempt to improve the resolution of the entire image cannot be achieved. On the other hand, using the information available from the entire image to improve the resolution of selected features or smaller regions does not violate principle of information content, and is therefore mathematically sounded. One technique that has been used for this application is the artificial neural networks (ANN) [26][27]. It has been shown that, by using appropriate ANNs, linewidths of narrow tracks 30x smaller than the resolution limit of the optical microscope could be measured with repeatability in the region of 1nm.

Although ANNs have been shown to be extremely powerful in terms of dimensional measurement, they have their shortcomings. Each ANN is trained to perform a certain operation. If used inappropriately, one would obtain wrong results. Hence the aim of this project to classify types of features, so that correct ANNs can be used. Some brief mention of the one used inappropriately techniques used as classifiers will be made.

We have chosen to take a different approach to this problem by attempting to directly measure parameters of objects below the classical resolution limit without trying to increase the system bandwidth. This technique is not providing an increase in resolution as the images obtained by the system are still diffraction limited. However the measurement range is increased to include the parameters of objects that are well below the conventional limits of the optical system thus providing a way to measure very small structure parameters optically. Various signal processing techniques, including principal component analysis (PCA) and Hough Transform (HT) will be used to extract and classify key object features. The dimensions of these

features are then measured, thus allowing object image of much greater details to be reconstructed. These features have various parameters associated with them such as width, length, height, separation and orientations. The flow chart of the whole system of the thesis is shown in Figure 1.

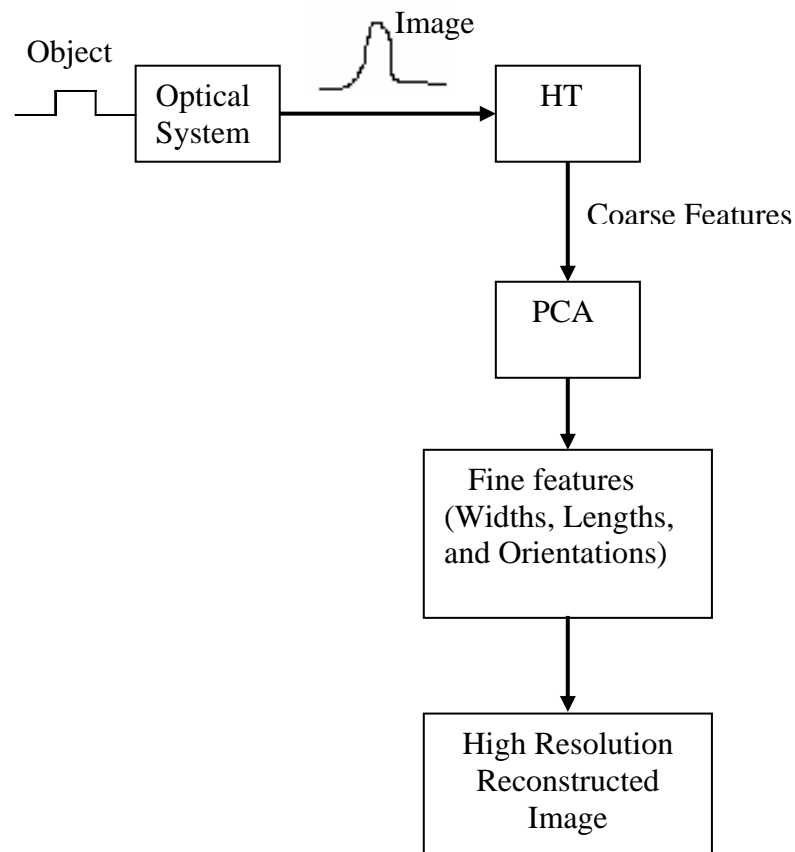


Figure 1.1 – The system process

From Figure 1 the image from the optical system is used as input to Hough transform (HT). The HT will classify and measure the main parameters of image such as number of lines, orientations, widths and lengths. At this process we obtain the coarse features structure of the whole image. Then individual feature will be passed

to perform the fine measurement by using PCA. After that the new image can be reconstructed with higher resolution based on the measured features.

1.3 Layout of thesis

This thesis is organised as described below:

Chapter 1 contained an introduction of the research project. Then we introduced a number of modern microscopy techniques that can achieve super resolution and the recent developments of microscopy were described in this chapter.

In chapter 2, we will begin with discussing the common resolution criteria and diffraction limit of a conventional light microscope. After that, prior research that aimed to increase the effective bandwidth of the system, or to extract high precision dimensional measurements, will be discussed. These techniques are software based and aim to extract information that is originally outside the system bandwidth from the image or profile of the sample. We will describe the mathematical basis that allows these techniques to work. We will also describe the measured profiles in term of their information contents, which will provide an upper value as to the ultimate resolution that can be achieved.

In chapter 3, we will describe the original HT and give some application examples. The modification and the application to our project will be described in the next chapter. We also introduce the basic mathematical concept of PCA and some example of its general applications.

In chapter 4, the modification and the application of Hough Transform to our project will be explained first. The details of the modified algorithm for detection of line segments in grey scale images directly will be discussed. The capability of the modified algorithm will be demonstrated through simulations. This will be followed by detailed description of the application of principal component analysis to our project.

In chapter 5, we will describe the extension of the PCA to two-dimension line segment images. Thus features of line segment such as lengths and shapes, which could not be measured using the line profiles, can now be determined. The use of PCA to measure these parameters from the line segment images will be described first. After that we will discuss the use of Hough transform. The purposes of HT are to provide initial information, mainly relating to location, orientation, shape, and rough dimensions of object, before applying 2D PCA to extract more accurate measurements.

Chapter 6 shows the experimental results to demonstrate the practical ability of the technique. The details of the experiment are carried out step-by-step.

In the last chapter, the general conclusions are presented. Future development and applications of PCA will be proposed.

Chapter 2

2 Literature review

The research project addressed in this thesis concerns the determination of dimensions of features that are too small for conventional optical techniques to measure precisely. The approach we take is, based on the profiles of the features, to determine the types of feature under examination (classification). Once this task is accomplished, signal processing techniques such as Hough transform (HT), principal component analysis (PCA) will be applied to the profiles in order to extract the dimension of this part. In this section, we will begin with discussing the common resolution criteria and diffraction limit of a conventional light microscope. After that, a number of prior researches that aim to increase the effective bandwidth of the system, or to extract high precision dimensional measurements, will be discussed. These techniques are software based and aim to extracting information that is originally outside the system bandwidth from the image or profile of the sample. We will describe the mathematical basis that allows these techniques to work. We will also describe the measured profiles in term of their information contents, which will provide an upper value as to the ultimate resolution that can be achieved.

Finally, a commercial microscope will be described. The reason for discussing such a system is that it can be used to provide highly reliable and repeatable data, both in amplitude and phase, for our technique. Such an instrument is essential if the technique described in this thesis is to be realised in practice.

2.1 Resolution Limit

The ability to resolve fine features is fundamental to a microscope. As technology progresses the demand to resolve ultra fine features has also increased. The application areas include medical science, biological science and nanotechnology industry. However, the optical microscopes have a finite resolution due to the nature of light. Waves diffract as they propagate. The light source, the medium and any obstacles in the beam path will affect the diffraction pattern. For example, if a plane wavefront is incident on a circular aperture, the intensity pattern observed in the optical far field is that of an Airy disc [28], given by

$$I = I_o \left[\frac{J_1(kNAr)}{kNAr} \right]^2 \quad (2-1)$$

Where I is the intensity distribution

J_1 is the first order Bessel function of the first kind

NA is the numerical aperture of the system

r is radial coordinate of the image plane and

$$k = \frac{2\pi}{\lambda} .$$

The numerical aperture of the system, NA , is defined as the ratio of the aperture radius to the distance between the aperture and the observation plane. Eq. (2-1) is sometime referred to as diffraction limited, in the sense that it is the smallest distribution that can be achieved under the stated arrangement, and with the assumptions that it is free from any aberrations. In an imaging system, the Airy disc

is also the image distribution for a point object, which is known as the point spread function (psf). Its non-zero size is due to the finite bandwidth of the optical system. To understand how it affects the system resolution, we consider imaging two point objects. If they are well separated, so would the corresponding images. As the two point objects move towards each other, the two Airy discs will start to overlap. The sum of the two peaks will appear as one and it will be difficult to tell whether one or two objects are responsible for the image.

Figure 2.1 shows the two individual images of the point sources and the actual image of the two point objects (dotted line). In this case the two point objects have not been resolved as only one peak is present in the image.

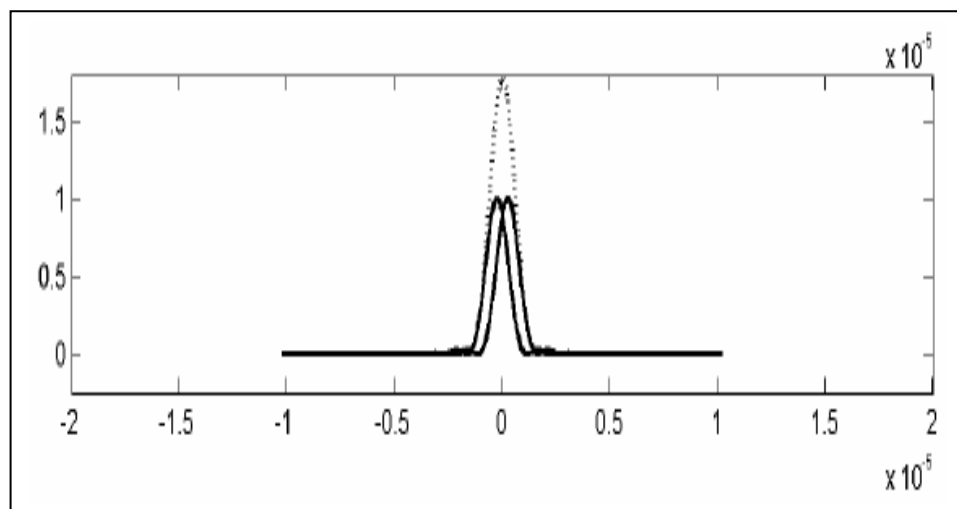


Figure 2.1. two point objects and corresponding image

Resolution criteria are subjective and the two used most commonly are proposed by Rayleigh and Sparrow.

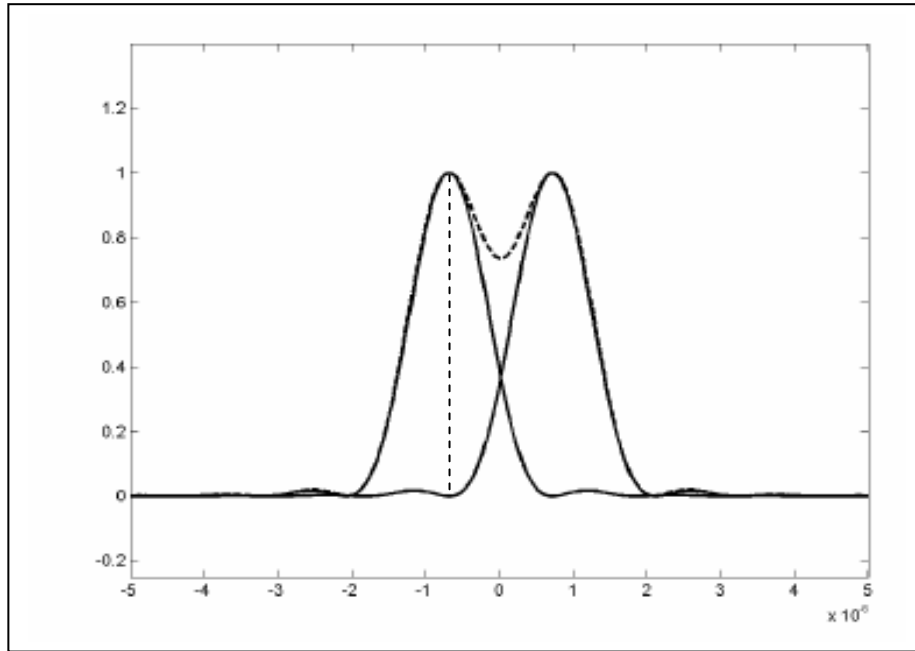


Figure 2.2. Rayleigh Criteria

The Rayleigh [29] criteria states that two point sources are regarded as just resolved when the principal diffraction maximum (the mainlobe) of one image coincides with the first minimum of the other as shown in Figure 2.2.

The mathematical expression for the optical system can be expressed by:

$$\Delta_R = \frac{0.61\lambda}{NA} \quad (2-2)$$

Where Δ_R is the object separation.

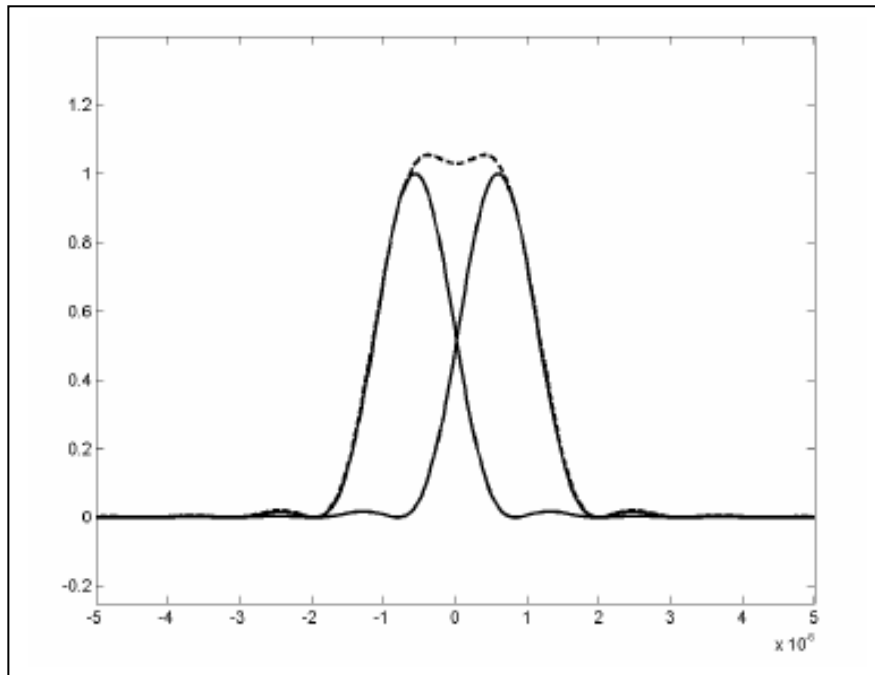


Figure 2.3. Sparrow criteria

The Sparrow [30] criteria define the resolution as the separation when the combined image no longer has a dip in the mid point between them as shown in Figure 2.3.

This can be written in equation as:

$$\Delta_s = \frac{0.47\lambda}{NA} \quad (2-3)$$

These values for the resolution of the system are possible assuming that the optical system is perfect in everyway, meaning that it should be free from aberrations. Practically, the resolution of the optical system also depends on the quality of the optical components used such as the light source, the signal to noise ratio and the detector in the imaging system.

It should be pointed out that the Sparrow criterion does not provide better resolution than Rayleigh, as both are definitions. From the above equations the resolution of the optical system can be increased if shorter wavelength is used or if the numerical aperture is increased. However, as the NA of the system is given by $\sin(\theta)$ where θ is the half acceptance angle of the objective lens, the system NA is limited to a maximum of 1 when operating in air. The practical range of usable wavelength for optical imaging is restricted to the visible spectrum. These limitations of conventional microscope give its resolution to be around 250 nm. This resolution is not satisfactory for many modern applications nowadays such as in biological research and nanotechnology research.

2.2 Spectral extension and information theory

In 1955 Toraldo di Francia published a paper showing that under some conditions two different objects would produce identical images[31] . This has implications for super resolving algorithms as without *a priori* information the correct object cannot be reconstructed. Harris[22] went on to relate the ambiguous image to its spectral components saying:

‘... objects can be distinguished one from another as long as the spatial frequency spectra of the two objects are not everywhere identical in the pass band of the optical system’.

He proposed that the fundamental limit due to diffraction on resolution must be because two or more objects could produce identical images. Harris then showed

that, providing certain conditions are satisfied, two different objects will never produce the same image. This condition is that the objects must be bounded, which in practise is usually the case for most imaging systems. This means that the limitation shown by Toraldo di Francia does not apply to imaging systems where the object is bounded. Harris' theory is underpinned by the fact that the Fourier transform of a bounded structure is an analytic function. Utilising the uniqueness theory and analytic continuation he showed that in theory arbitrary resolution is possible in a noiseless system.

2.2.1 Analytic continuation and the uniqueness theory.

An analytic function is a function that may be complex and that is infinitely differentiable. These functions have a special property [32][33], usually referred to as analytic continuation, which states that

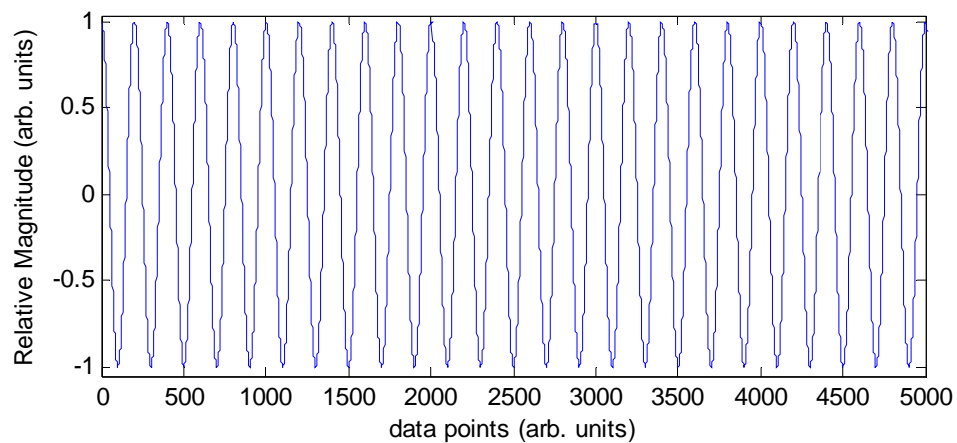
a function of a complex variable is determined throughout the entire z -plane from a knowledge of its properties within an arbitrary small region of analyticity

The uniqueness theorem states that

any two functions of a complex variable whose values coincide over an arbitrarily small region of analyticity must have identical values throughout their common region of analyticity and hence be identical.

These properties of analytic functions imply that if any part of an analytic function is known, then the entire function can also be determined as the known values can only belong to one specific function. It is therefore possible to reconstruct the spatial frequencies outside of the pass band of the optical system, based on the spectrum inside. An extension on an analytic function will always produce the same answer; the extensions are not ambiguous.

The influence of the bounded object on pass band of the optical system is discussed here. The example below shows a grating structure whose frequency is outside of the pass band of the optical system. If the object is infinite in extent as in Figure 2.4a, the spectrum is two delta functions at the grating frequency. If the frequency cut-off of the aperture of the system (shown in dashed line in Figure 2.4b) is below that of the object, no information will be in the pass band and as a consequence, no spectrum extension can take place. This demonstrates for an infinite object, the Fourier transform is not analytic.



(a)

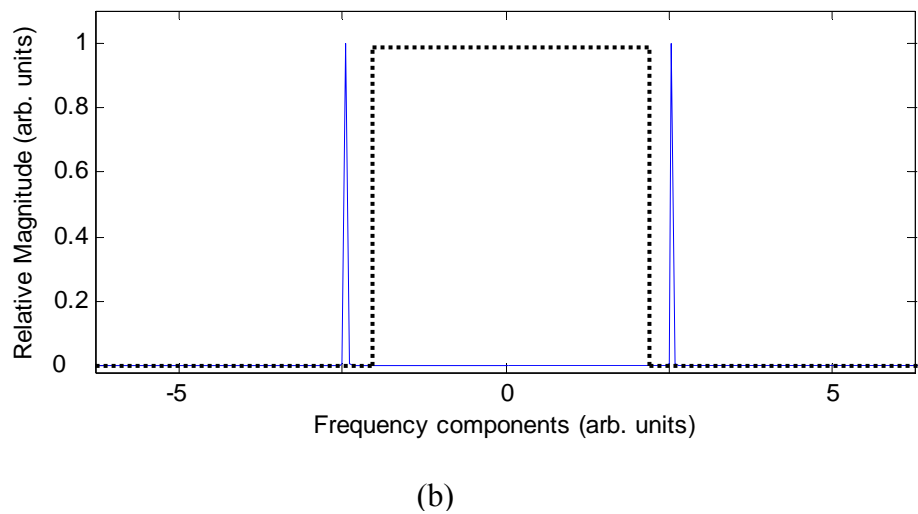


Figure 2.4 Grating structure (a) and spectrum (b)

If the object is now bounded as in Figure 2.5 then the spectrum of the grating is convolved with the spectrum of the truncating window. Instead of having two delta-functions, we now have two sinc functions centred at the locations of the original delta-functions. Although the peak of each of the sinc function is still outside the aperture of the system, certain information is leaked into the aperture, and it is this information that provides the basis for the eventual spectrum extension.

A sinc function can be expressed as a power series and by definition, is infinitely differentiable, hence analytic. Thus the information inside the pass band can be used to calculate the part of the spectrum outside by invoking the principle of analytic continuation.

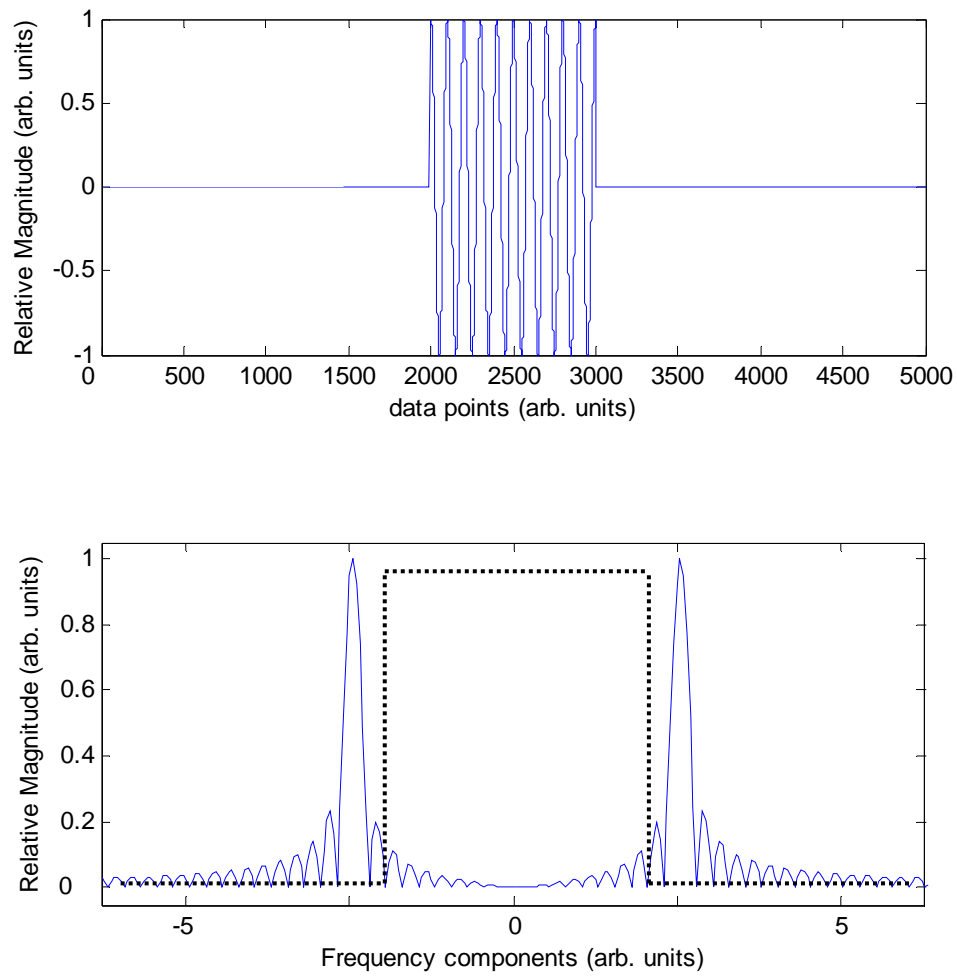


Figure 2.5 Grating structure of truncated object (top) and spectrum (bottom)

2.2.2 Information theory

The super resolution problem has also been considered from an information theory point of view. This is an attempt to establish the theoretical limits on extension and understand the influence of noise on different extension techniques. In 1966 Lukosz [34][35] and Toraldo di Francia [36] suggested an invariance theorem, which states that for an optical system the number of degrees of freedom is fixed. After that Cox and Sheppard [37][38] went on to develop this idea and took into account random

noise in the system and its effect on the resolution improvement that was possible. The information capacity equation developed is shown in Equation 2-4.

$$N = (2L_z B_z + 1)(2L_y B_y + 1)(2L_x B_x + 1)(2TB_T + 1) \log(1 + s/n) \quad (2-4)$$

Where N is the degrees of freedom, L_x L_y L_z are the extent of the field of view in the x,y,z directions, B_x B_y B_z are the spatial bandwidths in the x,y,z directions, B_T is the temporal bandwidth, T is the observation time, s is the signal level and n is the noise level.

This is a very useful equation as for any optical system it can be used to calculate the information capacity of the system and how this varies with signal to noise ratio. It should be pointed out that N is the theoretical maximum information capacity available from the system and so in practice the total information carried by the system may be much less. For example, the object being measured by the system may not vary in time, although the system maybe able to carry time information no actual object information is delivered by the system. This does however mean that if the object is known *a priori* to be restricted in anyway, additional information may be encoded onto the independent and unused parameters of the system.

Another way to understand this invariance theory is that, once an image is taken with a system, the total information content is fixed. By using mathematical or signal processing techniques, one can redistribute the weights among the different parameters, but the value N cannot be exceeded. Therefore if the spatial bandwidth

B_x is to be increased, there will be a corresponding decrease in the other terms in the equation to keep N fixed.

Super resolution in optical microscopy was considered by Cox and Sheppard [37][38]. They showed that the SNR of the super resolution image decreased as the spectrum was extended, if all other parameters were left the same. Using the above equations, one can calculate the SNR of the final image with an extended bandwidth. Because of the logarithm relationship in Eq. (2-4) , the noise in the extended image increases dramatically and unless the original image has an extremely high SNR then the extension yields either very poor extension or very poor SNR in the final image.

2.2.3 Analytic continuation by Taylor expansion

The function $f(x)$ can be extended outside of its known range by, for example, a Taylor series expansion, assuming that the function is known precisely over some arbitrary region centred at x_0 , the remaining function can be calculated as:

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \dots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0) \quad (2-5).$$

The function $f(x)$ is now known and if enough orders are used then this function is valid for all x and can be used to obtain sufficient values of $f(x)$ outside the known extent of the signal.

The success or otherwise of the extension depends on how well the starting function is known. If there is any noise involved then the extension will become less and less

accurate. This noise can come from inaccuracies in the measurement system but also from digitisation of the function, as using discrete levels introduces uncertainties in the actual value for each specific point. The differentiation in the Taylor expansion will also amplify noise. This will in practice greatly reduce the ability to perform an extension.

2.3 Applications of spectrum extension theory

2.3.1 Auto-Regressive Models

An Auto-Regressive (AR) model is a random process which is often used to model and predict various types of natural phenomena. It is also known as the maximum entropy model (MEM) in physics applications. This kind of model is one of a group of linear prediction formulas that attempt to predict an output of system based on the previous output, and is particularly suitable if the function concerned is oscillatory. It can be used to reconstruct the lost spectral components if some part of the spectrum is known. Thus the values of spectrum are outside of the bandwidth can be obtained point by point. This will always yield the extension (as the function is analytic and oscillatory) if there is no noise present in the system.

Definition:

The notation $AR(N)$ indicates autoregressive model of order N . The $AR(N)$ model is defined as

$$X_t = \sum_{i=1}^N c_i X_{t-i} + \varepsilon_t \quad 2-6$$

where X_t is the series under investigation, c_i are autoregressive coefficients and ε_t is white noise. Clearly, X_{t-i} denotes past values of the function.

In the absence of noise, Eq.(2-6) can be donated in the matrix form

$$Xc = x. \quad 2-7$$

Where X is the previous value and x is the current value of the function.

The AR coefficients are therefore given by

$$c = X^{-1}x \quad 2-8$$

if X^{-1} exists.

The difficulty of this method is that to determine the AR coefficients. The most common technique for deriving the coefficients is done by using Yule-Walker equations. Multiplying Eq.(2-7) by X' and then taking the expectation yield.

$$E(X'X)c = E(X'x). \quad 2-9$$

where E represents the expectation. By using the auto-covariance matrix R and vector r , Eq.(2-9) can be expressed as

$$Rc = r \quad 2-10$$

The solution is given by

$$c = R^{-1}r \quad 2-11$$

This is the least-squares solution of Eq.(2-7).

The model can be applied to the spectral expansion by dividing the known portion of spectrum into two sets. The first one is for X , hence the previous values are known. The second set is the actual current values of known points used for x . This allows the AR coefficients to be calculated. When all of the coefficients for the known spectrum are obtained, we can use these to calculate the spectrum outside the pass band point by point. The model will be updated after each new point of x is incorporated to the model and its coefficient is calculated.

The model looks simply in theory but the problem becomes more demanding in practice. This is because the way to determine coefficients of the model. Performing the inverse matrix in Eq.(2-8) is difficult as it is usually ill conditioned. In many practical cases the matrix X^{-1} is singular and its inverse does not exist. In addition the problem with performing inverse is that when round off and noise are involved. The small amounts of error in the elements lead to the large errors in the inverse.

This means calculating the inverse cannot be performed in the usual method. The technique used instead is the generalized inverse which is often calculated by the singular value decomposition (SVD). Performing SVD reduces the effect of noise and allows the inverse to be calculated efficiently.

The success in using an autoregressive model with singular value decomposition was shown in a paper by Minami et al [39]. They applied it to derive super resolution spectra of Fourier Transform Infrared (FT-IR) absorption data of benzene and cyclohexane. The result from their work showed an increase in spectral extension in the order of 8 times. However, this impressive extension is due to the fact that the function under consideration consisted of several sinusoidal oscillations of different frequencies. For a more complicated function, the method again yields very limited extension even when the presence of noise is small.

2.3.2 Sampling theorem in frequency domains [22]

In the frequency domain the sampling theorem states that an object function of width W which is in a range $\pm W/2$ has a spectrum that is exactly determined for all frequencies by specifying the value of the spectrum at discrete frequencies separated by the interval $1/W$, this series extends throughout the entire frequency domain. If we have an object that is bounded in X by $\pm X/2$, it can be described in the bounded region by the Fourier Transform pairs as :

$$N(x) = \sum_{-\infty}^{+\infty} G_n \exp[-i2\pi(nx/X)] \quad 2-12$$

and

$$G_n = \frac{1}{X} \int_{-X/2}^{X/2} N(x) \exp[i2\pi(nx/X)] \quad 2-13$$

From the sampling theorem the relationship between spectrum $G(f_x)$ and the series coefficients G_n can be obtained directly by substitution of Eq.(2-12) for $N(x)$ into Eq.(2-13) and then performing integration yields:

$$G(f_x) = X \sum_{-\infty}^{\infty} G_n \frac{\sin \left[\pi \left(\frac{n}{x} - f_x \right) X \right]}{\left[\pi \left(\frac{n}{x} - f_x \right) X \right]} \quad 2-14$$

This equation shows specifically the mathematical nature of the spectrum $G(f_x)$. With this approximation the summation in Eq.(2-14) becomes a finite series. It can be used as a basis for reconstructing the detail in diffraction image.

Applying the method to spectrum extension can be done by selecting a number of f_x equal to the number of unknown G_n 's and a set of simultaneous equation can be formed which can be solved for the G_n values. Once the G_n values are known, the object allows to be constructed by utilising Eq.(2-12).

The equation can be solved as follow. The Eq.(2-14) can be written in the matrix form as

$$G_k = SG \quad 2-15$$

Where G_k is the $(2N+1)$ column vector of known frequency components. G is the $(2M+1)$ column vector of unknown coefficients of the G_n 's. S is the $(2N+1) \times (2M+1)$ *sinc* function matrix.

The unknown coefficients contain information beyond the cut-off of system can be obtained from the known spectrum and *sinc* function matrix S , thus the Eq.(2-15) can be rearranged as

$$G = S^{-1} G_k \quad 2-16$$

Harris demonstrated this technique by applying it to an object consisting of two incoherent point sources and showed a large increase in resolution. However this approach is only reliable in the absence of noise. Once a system has noise involved the uniqueness theorem no longer applies, as the spectrum of function with random noise is not necessarily analytic. Therefore the extension limitation of the technique relies on the noise level of the system. Moreover, the precision requirement of the technique is that the interval X must completely encompass the object and have to be minimised by making the interval as small as possible.

2.4 Artificial Neural Networks

An artificial neural network (ANN) is a mathematical model or computational unit, which is inspired by the structure and/or functional aspect of biological neural

networks. In an artificial neural network, simple artificial nodes or units are connected together to form a network of nodes mimicking the biological neural network, hence the term “artificial neural network”. As its biological predecessor, thus an ANN is an adaptive system. By adaptive, it means that each parameter is changed during its operation and it is deployed for solving the problem in matter. This process is called training phase. An ANN is developed with systematic step-by-step procedure which optimises a criterion commonly known as the learning rule. The system is a structure which receives input data, processes the data and provides output. Commonly, the input consists of data array which can be any kind of data and can be represented in array such as data from image files and wave sound. Once an input data presents to the ANN, and corresponding desired or target response is set at the output, an error is calculated from the difference of the target and the real system output. The error is then fed back to system which makes all adjustments to parameters in systematic fashion (known as learning rule). This process is repeated until the desired output is met. It is important to notice that the performance of the system hinges heavily on data, hence this is why the input data should pre-process with the third party algorithms such as digital signal processing (DSP) or Fourier Transform (FT) before presented to the network.

2.4.1 The ANN model

Artificial neural networks came about after McCulloch and Pitts introduced a set of simplified neurons in 1943. These neurons were represented as models of biological networks into conceptual components for circuits that could perform computational tasks. The basic model of the artificial neuron is founded upon the functionality of

the biological neuron. By definition, “Neurons are basic signalling units of the nervous system of a living being in which each neuron is a discrete cell whose several processes are from its cell body”.

Once modelling an artificial functional model from the biological neuron, we must take into account three basic components. First off all, the synapses of the biological neuron are modelled as weights W_n . These weights are multiplied by the input data X_n . The synapse is the one which interconnects the neural network and gives the strength of the connection. The following components of the model represent the cell body which is modelled by the sum and activation function $f()$. The node calculates the weighted sum of its input. The output from the node is determined by the activation function. This will be discussed some more detail in the next section. Finally, the output y represents axon of the neuron. This simple artificial model shows in Figure 2.6.

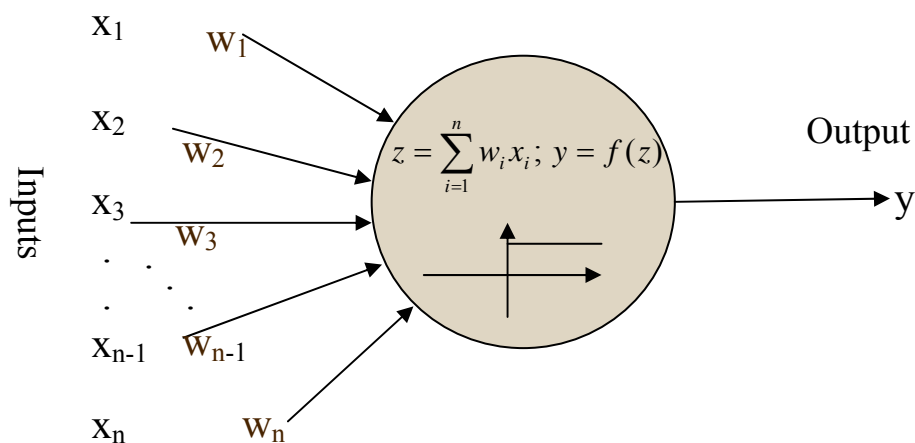
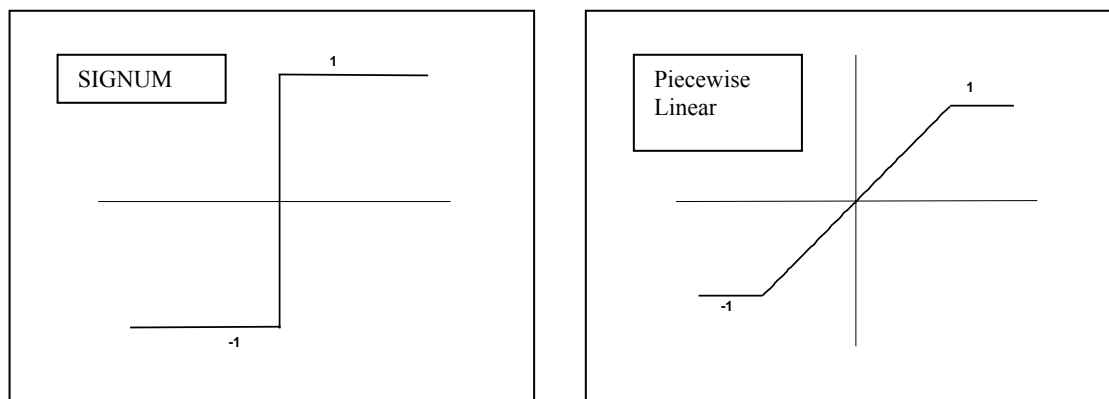


Fig. 2.6 An artificial neuron model

2.4.2 Activation Functions

The activation functions are the function used to calculate output and control the amplitude of the output level of the node based on the values from the sum of weighted input to that node. They play important role in the training/learning processes. For example if a linear activation function were used, the network can only work with the linear system. This is because the linear combination of linear function is still a linear function. If a nonlinear activation function were used, any nonlinear function can be approximated. In addition they act as a squashing function, such that the output of a node in an artificial neural network is between certain values (usually 0 and 1, or -1 and 1). Activation function with discontinuities are usually not used as it is then difficult to train the network, as the most common training methods rely on a continuous derivative of activation function such as delta rule. In general, there are many types of activation functions used. Some of them are shown in Figure 2.7.



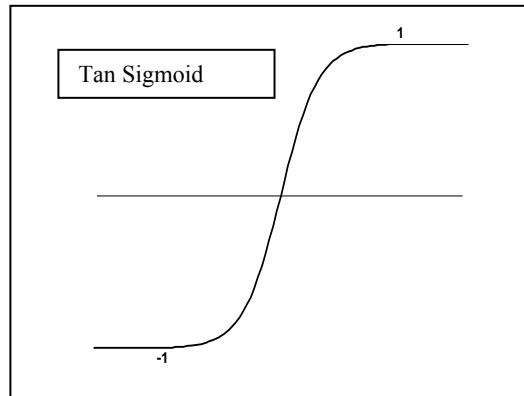


Figure 2.7. Activation function

2.4.3 Network topographies and applications

An artificial neural network consists of combinations of many neurons to form a network. The topography of the network, pattern of connections between the units and the propagation of data, determines its suitability for different applications.

There are several categories of networks classified by their topography, these are:

- **Feed-forward neural networks-** where the data flow from input to output units is strictly feedforward. The data processing can extend over multiple units (layers) but no feedback connections are present. Each successive layer is connected to the next either partly or completely. An example diagram of network is shown in Figure 2.8. Examples of feed-forward neural networks are the Perceptron and Adaline. These types of network are normally applied to data classification, pattern recognition or functional interpolation.

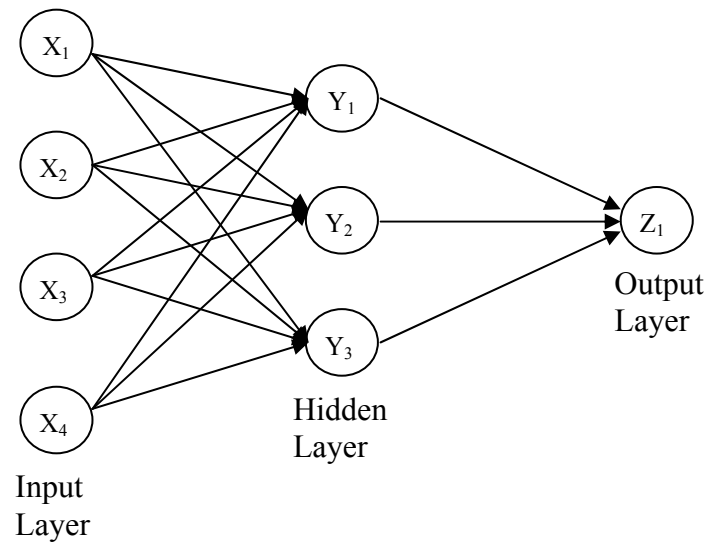


Figure 2.8 Fully connected feed-forward network with one hidden layer and one output layer

- Recurrent neural networks-** networks have feedback loops between the layers. Figure 2.9 shows a three-layer network with the addition of a set of “context units”. There are connections from the hidden layer to these context units fixed with a weight of one [70]. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the neural network will evolve to a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behaviour constitutes the output of the neural network. They are typically used for associative memory, noise filtering and content addressable memory [40].

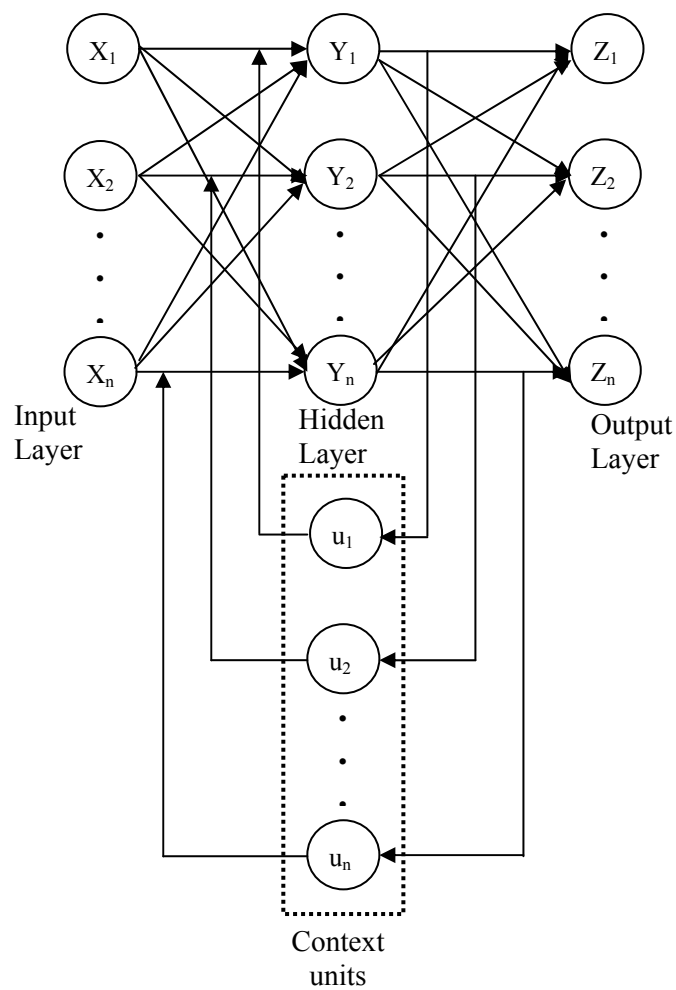


Figure 2.9 Simple recurrent network

- **Competitive neural networks** – these tend to be self-organising and are used for analysis of topological features and cluster template formation [41]. Competitive learning is implemented with network that contains a hidden layer which is commonly called “competitive layer” [71]. For every input vector, the competitive neurons competes each other for the winner neuron. The winner neural m set its output to one and all the other competitive neurons set their output to zero. The network diagram is shown in Figure 2.10.

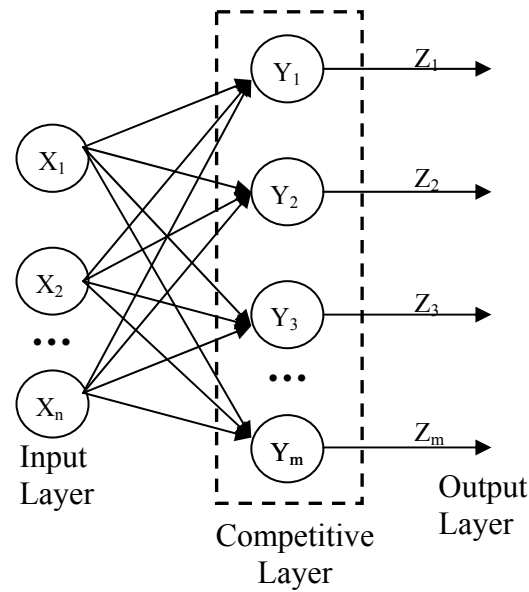


Figure 2.10 Competitive network architecture

2.4.4 Training network

This process is required to generate the weight values for the network to perform the desired task. The way is to train the neural network by feeding it with teaching patterns and letting it changes its weights according to some learning rule. Training the network can be either supervised or unsupervised depending on the network type.

- **Supervised learning** or Associative learning in which the network is trained by providing it with input and matching output (target) patterns. The network then calculates its output and adjusts the weights to reduce the overall error after each training pattern pass through it. This process iterates until some predefined stopping criteria is reached.

- **Unsupervised learning** or Self-organisation in which an output unit, no target given, is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input data and it must develop its own representation of the input stimuli. After the training process is completed the weights are usually fixed for normal operation and network stops learning.

2.4.5 Application to optical measurement

The application of ANN to super-resolution, measuring line width below optical resolution limit, is shown in papers by Smith *et al* [26][27]. They used ANN to extract parameters, track width and track separation, from profiles of chrome and silicon images sample. The ANN used in their work was feed-forward network with eight input values, five hidden nodes and one output node. The network was trained by supervised learning with backpropagation algorithm. An activation function was tansigmoid function. In this case the output of the network is either track width or track separation. Their experiment used 0.3 NA objective lens and laser wavelength of 633 nm, this makes the PSF, at the sample surface, in the order of $2.6 \mu m$. The input data, surface profiles, were obtained from an ultra stable common path scanning optical interferometer. This data is pre-processed before it can be used as the ANN input such that the profiles are differentiated and then discrete Fourier transform is applied to yield the spectrum of the differentiated profiles. The spectrum amplitude is then sampled at eight equally spaced locations inside the spectrum bandwidth. The eight sampled values are used as the input to the ANN. The result

demonstrates that the technique can measure track width down to 60 nm with repeatability better than 2 nm. This means their system could measure the track width less than 40 times of the PSF. The technique had been also applied to classify samples of various dimensions. The system showed perfect classification although the samples are very similar to each other. For example, it could distinguish a double track of width 500 nm (each track) and separation of 50 nm, to a single track of 1050 nm width.

Although the technique has been demonstrated to have a great potential to measure and classify the features of objects below the conventional diffraction limit of conventional optical microscope, one network is only valid and reliable for a certain range of trained set of input data. The system will produce enormous error for out of range data. It is not easy to determine range coverage. This means more than one network will be required in order to cover the measurement. As a result the methods to select a suitable range and network are necessary.

2.5 Commercial microscope

A commercial microscope that could be used to provide line image segments will now be discussed.

ZYGO NewviewTM 700s [42]

The Newview 700 is a commercial microscope provided by Zygo for the applications of measurement, three-dimensions and scanning profiles etc. Like most

optical microscopes it provides fast, non-contact measurements. It uses integrated long life white light LED with computer controlled light level as the light source. The interferometer works as a Mirau or Michelson type and uses objective lenses with an internal beam splitter and reference path that matches the optical path length. The system provides sub-nanometre z resolution $<0.1\text{nm}$ with sample step heights up to 0.15mm . The lateral resolution and field of view is dependent upon which objective is used from the objective turret. The best lateral resolution is quoted as 360nm . The step height accuracy is quoted as $\leq 75\%$ and the repeatability is better than 0.1% . The system comes with a software package ZYGO Metro running under Microsoft Windows XP. This system is capable of providing many different types of measurements on various types of surface and samples.

Olympus – Confocal Laser Scanning Microscope – LEXT [43]

LEXT is a commercial optical system suitable for many imaging applications. It has several different operation modes such as confocal, darkfield and DIC Nomarski all as scanning modes using laser illumination. It can also provide colour images by illumination by a white light halogen source for wide field images. The system has a large sample area and like most microscopes it does not require special sample preparation. The repeatability is important if the system is to be used for linewidth measurements, depth measurement and surface roughness etc. The repeatability of the system is quoted as $0.002L\mu\text{m}$ (L =measurement length). The light source used has a wavelength of 408nm . The lateral resolution of the microscope is quoted as $0.12\mu\text{m}$. The reliability of the measured data is traceable to international standards set out by Physikalisch-Technische Bundesanstalt (PTB), Japan Quality Assurance

organization and the United Kingdom Accreditation Service (National Physical Laboratory).

Chapter 3

3 Hough Transform and PCA

In this project, two methods are used for processing of the data, and they are the Hough Transform (HT) and Principal Components Analysis (PCA). The HT is used extensively to detect position and orientation of straight line segments in an image area of interest. It is a transformation from image space to another parameter space, sometime known as the Hough space. The intention is to reduce a feature to a point in parameter space which eases the processing for detection purpose. Its main application is for detecting the edge pixels in a binary image. For a gray scale image, therefore, it must be pre-processed using an edge detection technique. In our project, we modify the original Hough transform and apply the new algorithm to detect the positions and orientations of line segments in grey scale images. These line segments may be in the form of isolated segments, or multiple segments not in contact with each other, or multiple segments that intersect each other. Figure 3.1 shows examples of the three types of features.

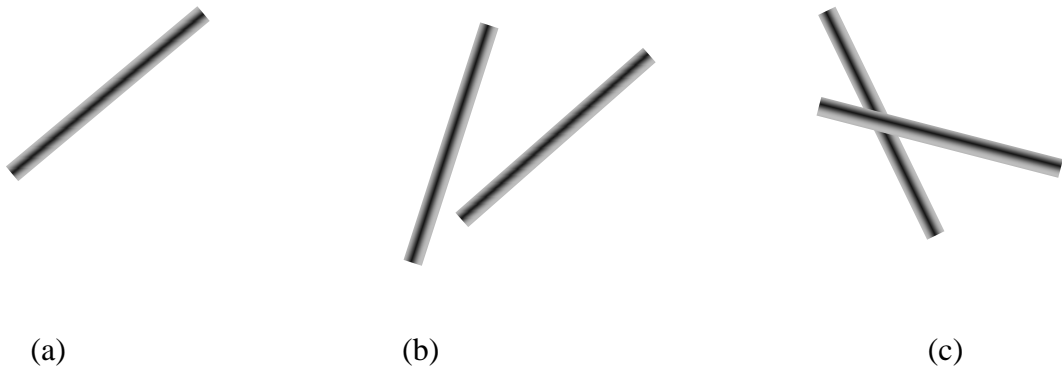


Figure 3.1 The form of line segments: a) single line segment b) multiple segments not in contact each other and c) multiple segments intersect each other

The PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated data into a set of values of uncorrelated variables called principal components. It is a way of identifying pattern in data and expressing the data in such a way as to highlight their similarities and differences. We adopt this technique to identify features in an image, by comparing them with templates generated with known objects. For our application, will apply the Hough transform first to determine the orientation and location of the features, which is then followed by the PCA to classify the features.

In this chapter, we will describe the original HT and give some application examples. The modification and the application to our project will be described in next chapter. We also introduce the basic mathematical concept of PCA and some of its general applications. Its application to our project will be detailed in the next chapter.

3.1 Hough Transform

The Hough Transform is a technique to detect features of a particular shape that can be parameterised such as straight line segments and circles in binary images. It was proposed by Paul Hough in 1962 [44]. The method provides a robust technique to identify the parameters of straight line edges in an image space. The technique converts the original problem of finding a straight line that passes through as many points as possible to one of computing points of intersection in the parameter space. The idea is to map data, object pixels in image space, into parameter space by a voting procedure, from which object candidates are obtained as local maxima in a so-called accumulator that is explicitly constructed by the algorithm for computing the

Hough transform. It can be applied to many computer vision, image analysis and digital signal processing applications as most images contain features boundaries which can be described by regular curves, although in the classical Hough transform it is restricted to line detection in binary image.

However, the original paper had lain dormant until 1972, when Dura and Hart [45] proposed an innovative approach to detect lines. The concept of Hough transform was used in their technique for straight lines and curves detection. The procedure involved grouping of geometric parameters of a sector of collinear edge pixels. For each edge pixel in an image space, the parameters of the corresponding normal parameterisations are mapped to Hough space. The technique was an extended version of Hough transform that can be applied to identify arbitrary shapes, commonly circles and ellipse.

Ballard [46] developed a method that deals with curves by using Hough transform. The technique exploits the duality between the points on the curve and the parameters of that curve. The method can detect the arbitrary shapes in both analytic and non-analytic curves. Since this technique has been presented through the journal article titled “Generalizing the Hough transform to detect arbitrary shapes”, this transform has been a popular technique in the computer vision community.

John Princen et al [47] have proposed a hierarchical technique to line detection based on Hough transform. The technique is an efficient method for finding lines and edge maps. The proposed algorithm is based on a pyramid structure. For each layer of pyramid, the whole image is divided into a number of sub-images. At the bottom

level of the pyramid, line segments are detected by applying Hough transform to the sub-images. This hierarchical approach is a well known technique for extracting lines from mechanical parts and engineering drawings.

Soo Chang Pei and Ji-Hwei Horng [64] have proposed a new method for detecting circular arcs. The technique uses Hough transform to detect circular arcs by using the centre locations and the radius as parameters. The peak value in parameter space of Hough transform indicates the existence of a circular arc of particular parameters in the image, whose values are given by the coordinates of the peak. The arc points are separated from the remaining edge points of the images, thus the arc length and the mid-point can be determined by analysing the data in the region near the peak location in the Hough space. The proposed algorithm can be used to detect the existence of multiple circular arcs in a noisy image.

3.2 Representation of lines into Parameter Space

The simplest case of the Hough transform is the linear transform for detecting straight lines. In the Hough transform, the idea is to consider the characteristics of straight line not as image points in x - y space, but in terms of its parameters (hence the term parameter space) here the gradient m of the line and its y - intercept c . Each line is mapped from the image space to the parameter space. The intention is to reduce the each line to a point in the parameter space which eases the processing for detection purpose. By doing this each pixel in the line contributes to a transformation process by “voting” for the lines. This results in an accumulation of votes in the parameter space. After the transformation, we search for the pixel that has the peak value, or the

highest number of vote. The locations of this pixel thus represent a line, in terms of its gradient and intercept, in the image space. A line in the image plane is therefore reduced to one point in the parameter space.

In the image space (x-y plane), a line has a unique value of slope (m) and intercept (c). The equation of a straight line is given in Equation (3-1) and can be graphically plotted for each ordered pair of image points (x,y).

$$y = mx + c \quad (3-1)$$

The two parameters, m and c , are used in the parameter space to represent the straight line. Therefore each set of collinear points are represented by (m,c) in the parameter space, or conversely each straight line in the image space contributes to a unique (m,c) value. For example the two lines in figure 3.2 will result in two distinct sets of (m, c) values, and each will be represented by a point in the parameter space. In the parameter space, c and m are linked using equation

$$c = y - mx \quad (3-2)$$

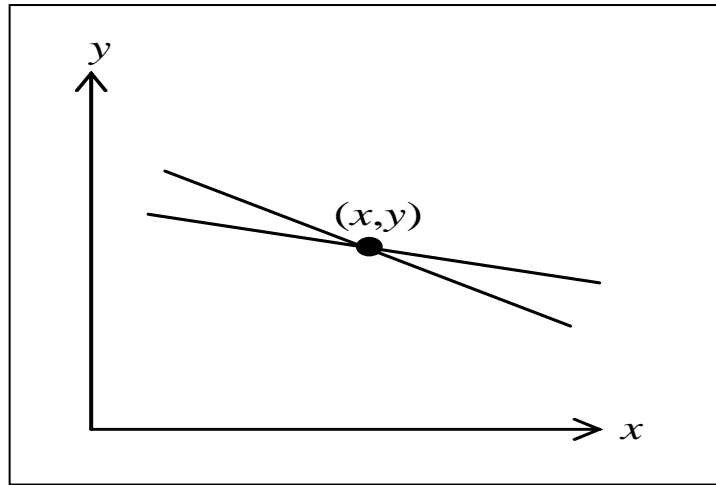


Figure 3.2. Lines through a point in x-y plane

Unfortunately, both the slope m and the intercept c are unbounded and cannot represent vertical lines. This is a severe drawback of the technique. This problem can be solved by using different parameters instead of slope and intercept. A suitable set of parameters are the normal distance ρ from the line to the origin and the angle θ which is defined as the angle of the normal of the line passing through the origin. These two parameters are shown in Figure 3.3. With the new parameters, a vertical line is given by $(\rho, \theta) = (d, 0^\circ)$ where d is the distance. The transformation is described by eq. (3.3) and is shown in figure 3.3.

$$\rho = x \cos \theta + y \sin \theta \quad (3-3)$$

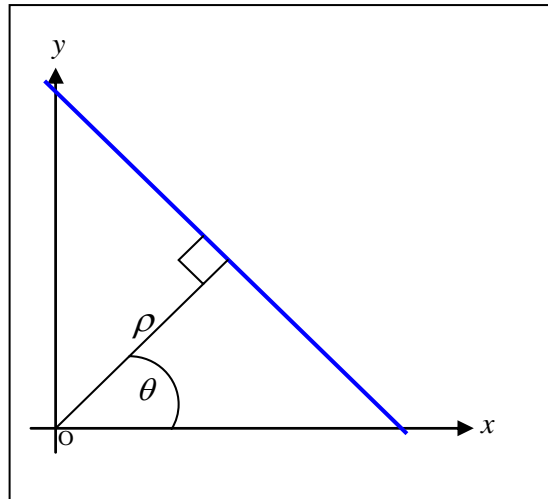


Figure 3.3 Representation of a line using parameter ρ and θ

The parameter space for straight lines has two dimensions, $\theta (\theta \in [-90, 90])$ and $\rho (\rho \in R)$ where R is some real number. A line is represented by a single point in the parameter space, corresponding to a unique set of parameters (θ, ρ) .

To perform the transformation, we consider the straight line in figure 3.4 (a). A point P on the line is transformed into a sinusoidal curve in the parameter space, according to eq 3-3, and is shown in figure 3.4 (b).

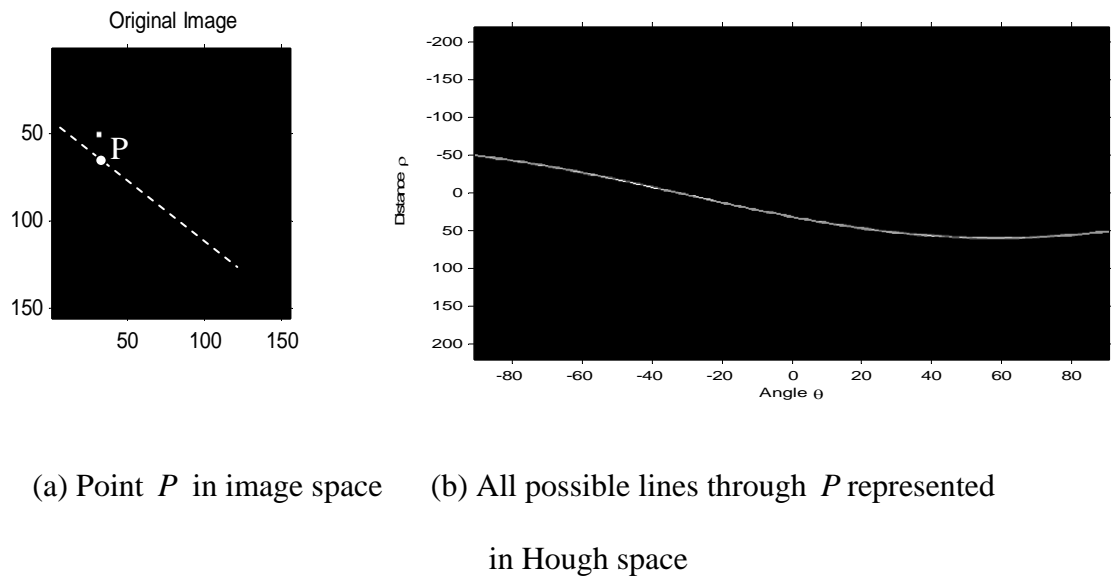


Figure 3.4 Transformation of point P to the line in the Hough space. The line in Hough space represents all possible lines pass through P

A different point on the straight line will result in a different sinusoidal curve. Indeed every point on the line will produce a different curve, but crucially they intersect at one point only. The coordinates of this point of intersection at the parameter space correspond to slope and intersect of the straight line in the x-y plane. The line-to-point mapping is illustrated in Figure 3.5. The simulation of the whole line transforming to parameter space will be described in the next section.

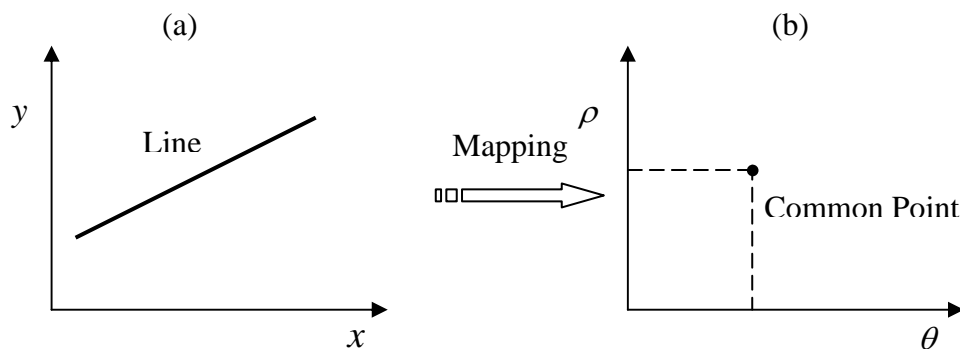


Figure 3.5. Mapping of a unique line in image space to the Hough space

3.3 Classical Hough Transform Algorithm for line detection

Following the discussion above, we now can describe the algorithm for detecting lines in image space. The steps are as follows:

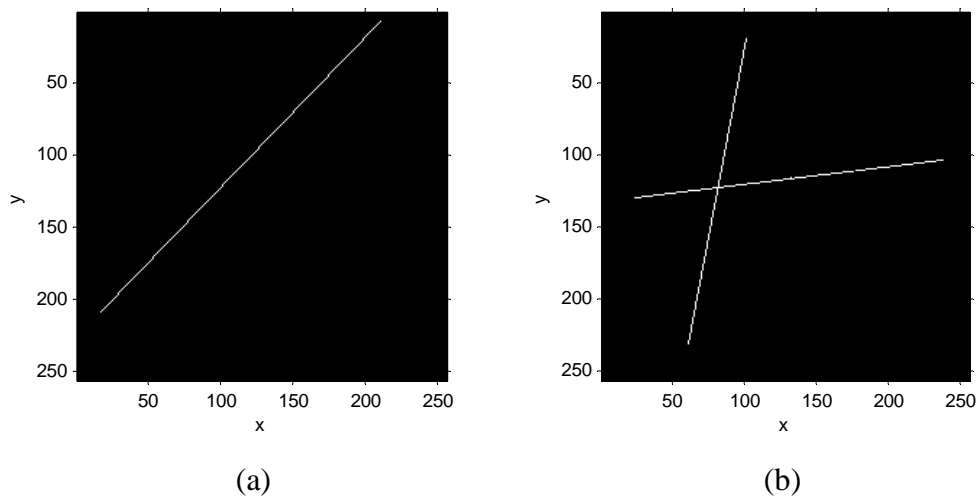
1. Create the parameter (ρ, θ) space as a two-dimensional matrix (figure 3.5b). This matrix is known as the Accumulator Array (AA). The quantisation levels along the two dimensions are user specified.
2. Apply a suitable edge detection method to the input image.
3. Initialise the matrix AA to zero.
4. Each edge point in the image space is used to create a sinusoidal curve, as described in eq. 3.3, in the parameter space (figure 3.4). The elements in the matrix AA, underneath the curve, will have their values updated by 1 as given by eq 3.4:

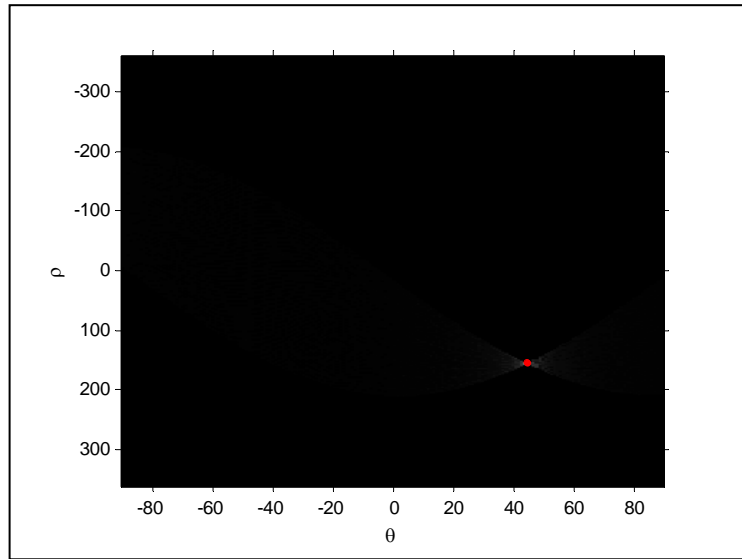
$$HT(\rho, \theta) = HT(\rho, \theta) + 1 \quad 3.4$$

5. Repeating step 4 for all the edge points would result in a vote matrix, showing the frequencies of the edge points for various (ρ, θ) values.

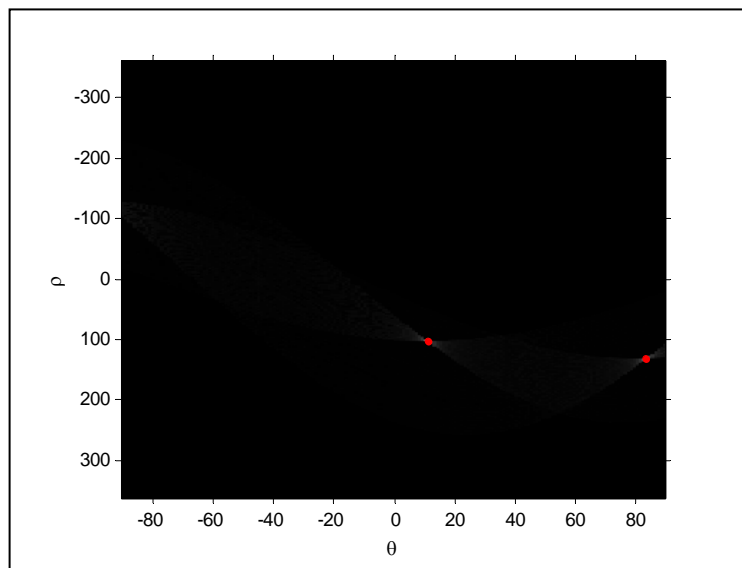
6. Interpretation of matrix AA would yield the line segments. The interpretation is done by thresholding and possibly other constraints where only elements with votes greater than a certain value are taken. These elements correspond to lines in the original image and the value of the element provides a measure of the number of points, or the length of the line.

Figure 3.6 shows a simulation of the Hough transform for lines detection. The original images are a single line (a) and two lines (b). The results of this transform were stored in a matrix AA. The value contained in each element represents the number of curves through that particular point. High element values are rendered red, as shown in Figure 3.6 (c) and (d). In Figure 3.6 (d), the two distinct red spots signify two lines in the original image. From the positions of these spots, the slopes and intersects of the two lines in the image space can be determined.





(c)



(d)

Figure 3.6 Lines detection using Hough transform: a) single line image b) double lines image c) Hough space correspond to (a) and d) Hough space correspond to (b)

The classical Hough transform can detect the presence of lines, returning the parameters ρ and θ . For our applications, lines would appear in the image space with finite lengths, different locations, non-zero widths, and possibly varying intensities along the lengths. The classical HT is not capable of measuring these different parameters. In the next chapter, we will describe algorithms we have generated. When combined with the classical HT, we can extract relevant parameters that can better describe images of real features.

3.4 Principal Component Analysis

Principal Component Analysis (PCA) is probably the most well known multivariate statistical technique and is used in almost all scientific disciplines. It is also likely to be the oldest multivariate technique. It was invented in 1901 by Pearson [48] but its modern formulation was due to Hotelling [49] who also coined the term *principal component*. The PCA is a mathematical procedure that is used to analyse a data set representing observations described by several dependent variables which are, in general, inter-correlated. Its goal is to extract the important information from the data and to express this information as a set of new orthogonal variables called principal components. It also represents the pattern of similarity and difference between two groups of data, and display them as points in maps [50][51][52]. This last point is the aspect of PCA that is used in our research. The PCA is a classical technique which can apply to linear domain thus data obtained from linear systems, such as signal processing, image processing, system and control theory, may be subjected to principal component analysis [53]. Furthermore, PCA is used to reduce the dimensionality of the data space (observed variables) to smaller intrinsic

dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables.

The following are some of the applications of the PCA. The Principal Component Analysis is one of the most successful techniques that have been used in image recognition and feature extraction.

Turk and Pentland [54] used PCA technique for face recognition. Their technique was to transform face images into a small set of characteristic feature images. The main idea of the principal component analysis in their approach is to find the vectors which best account for the distribution of face images within the entire image space. These vectors define the subspace of the face images and are called *face space*. The original image and its projection onto face space is shown in Figure 3.7.



Figure 3.7 An original face image and its projection onto the face space defined by the eigenfaces [54].

Each vector is of length N^2 , represents an $N \times N$ image, and is a linear combination of the original face images. These vectors are also the eigenvectors of the covariance matrix corresponding to the original face images, and are often referred to as the *eigenfaces*. Each face image in the original image space can be represented exactly in terms of a linear combination of the eigenfaces. The steps for face recognition can be summarised as follow:

1. Populate the image space by acquiring an initial set of face images, thus forming the training set.
2. Apply PCA to the training set to yield the eigenfaces, keep only M images which have the highest eigenvalues. These M images define the face space.
3. When a new face image is presented to the system, calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.
4. Determine if the input image belongs to the image space by checking the values of the projected weights from 3) above.
5. If the input face belongs to the image space, it will be classified further by comparing it with individual face images in order to determine if it is a known person.

The experiment of this system achieved approximately 96% correct classification and recognition. The eigenface approach provides a practical solution that is suitable for the problem of face recognition. It is fast, relatively simple and has been shown to work well in a constrained environment. However, the performance of the system

degrades with the light condition and the orientation variation of head. Moreover, the performance accuracy decreases dramatically with the change of head size.

Wei Qu et al [55] have proposed a robust segmentation approach for noisy medical images, X-ray and ultrasound images, using PCA model based particle filtering. The technique exploits the prior clinical knowledge of the desired object's shape information through the PCA model. They used the two nodes of the PCA to represent the principle component vectors for object region and background surrounding that object. Their experiment showed that the approach can achieve very robust segmentation and extraction results, even though the images may have much background noise, clutter, shadow and low contrast.

Mishra and Mulgrew [56] presented the use of PCA on radar images for automatic target recognition (ATR). The PCA is firstly applied to the training data set to reduce the number of observed variables, extract data as features, and then use as the classifier. Their approach was validated by using the synthetic aperture radar (SAR) images. The technique showed great promise for classification. Moreover, it is also a strong candidate for any real time ATR system, given its performance and computing speed.

Bajwa et al [57] used the principal component analysis to classify different types of single-layered and multi-layered cloud images for forecasting weather applications. The system reads features of gray-scale images to create an image space, training images. This image space is then used for classification of cloud types (clear sky, low-level, mid-level and high-level clouds). In testing process, a new cloud image is

classified by comparing it with the specified image space using the PCA algorithm. By using this technique, they could achieve high accuracy classification up to 90%.

3.5 Mathematics of PCA

Mathematically, PCA depends upon the eigen-decomposition of positive semi-definite matrices formed by the input data set. If the latter is ill conditioned or not square, singular value decomposition (SVD) is used to find the eigenvalues [58]. This section describes some background mathematical that will explain the process of Principal Components Analysis. The topics cover matrix algebra such as eigenvectors and eigenvalues which are the important properties of matrices and are fundamental to PCA. After that the steps needed to perform principal component analysis on a set of data will be explained.

3.5.1 Covariance Matrix

The covariance is a statistical analysis which is used to measure the relationship between data along different dimensions in a data set. For example, it is used to find how much the dimensions vary from the mean with respect to each other. Covariance is always measured between two dimensions. If a covariance is calculated from one dimension and itself, it is a variance of that data set. If we have a data set with more than two dimensions, say a n -dimensional data, there is more than one covariance measurement that can be calculated. For a n -dimensional data set, the number of different covariance values can be determined by

$$No.Dif.Cov = \frac{n!}{(n-2)! \times 2} \quad (3-4).$$

For a two-dimension data set (X and Y), the covariance value can be calculated as in Equation 3-5.

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \quad (3-5)$$

Where \bar{X} is the mean value of data X and \bar{Y} is the mean value of data Y .

If we have n -dimensional data set, it is useful to manage all the possible covariance values between all the different dimensions by arranged them into a matrix, n rows and n columns matrix. Each entry in the matrix is the result of calculating the covariance between any two dimensions, including the variances of the data contained within the individual dimensions.

Therefore the covariance matrix for a set of data with n dimensions is:

$$C = \begin{bmatrix} \text{cov}(D_1, D_1) & \text{cov}(D_1, D_2) & \cdot & \cdot & \text{cov}(D_1, D_n) \\ \text{cov}(D_2, D_1) & \text{cov}(D_2, D_2) & \cdot & \cdot & \text{cov}(D_2, D_n) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \text{cov}(D_n, D_1) & \text{cov}(D_n, D_2) & \cdot & \cdot & \text{cov}(D_n, D_n) \end{bmatrix}_{n \times n} \quad (3-6).$$

Where D_i is the i th dimension.

Note that the covariance values along the main diagonal of the matrix are the variances of the data along different dimensions. The other point is that $\text{cov}(D_i, D_j) = \text{cov}(D_j, D_i)$ since multiplication is commutative.

3.5.2 Eigenvectors and Eigenvalues

Eigenvectors and eigenvalues are quantities associated with square matrices, and can be obtained by using eigen-decomposition of the matrix. Even though eigen-decomposition may not be applied to all matrices, for example matrices that are non-square or contain singularities, it can always be applied to the correlation, covariance, or cross-product matrices formed using the original data. The eigen-decompositions of these matrices are important as they provide useful information such as the maximum (or minimum) variance among the original data. Specifically PCA is obtained by applying the eigen-decomposition to the covariance matrix of the input data.

There are several ways to define eigenvectors and eigenvalues, the most common approach defines an eigenvector of the matrix \mathbf{A} as a vector \mathbf{u} that satisfies the following equation:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \tag{3-7}$$

where λ is a scalar called the eigenvalue of \mathbf{A} corresponding to the eigenvector \mathbf{u} .

The Equation (3-7) is called eigenvalue equation or eigenvalue problem.

The vector \mathbf{u} has the property that its direction is not changed by any linear transformation of \mathbf{A} but is only scaled by a factor of λ , corresponding to a change in its magnitude. Most vectors will not satisfy Equation (3-7), a typical vector will change direction when it is multiplied by \mathbf{A} , so that $\mathbf{A}\mathbf{u}$ is not a scalar multiple of \mathbf{u} . This means that only certain special vectors \mathbf{u} are eigenvectors, and only certain special scalars λ are eigenvalues of \mathbf{A} .

To compute the eigenvectors of a matrix, it first computes its eigenvalues. When the eigenvalues are known, it is relatively easy to compute the corresponding eigenvectors. Equation (3-7) can be rewritten as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \quad (3-8)$$

where \mathbf{I} is the identity matrix.

If the quantity inside the brackets has an inverse, then both sides of Equation (3-8) can be left multiplied by the inverse to obtain the trivial solution: $\mathbf{u} = \mathbf{0}$. Thus we require there to be no inverse, and from linear algebra the determinant should be equal to zero:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \quad (3-9)$$

This equation is called the *characteristic equation* of \mathbf{A} . This characteristic equation will have N_λ distinct solution, eigenvalues, where $1 \leq N_\lambda \leq N$ and N is the number of dimensions of vector space which \mathbf{u} contained.

Based on the eigenvalues, the corresponding eigenvectors \mathbf{u}_i can be obtained by solving Equation (3-7) for each eigenvalue λ_i .

Usually, the eigenvectors of \mathbf{A} are arranged in a matrix (denoted \mathbf{U}). Each column of \mathbf{U} is an eigenvector of \mathbf{A} . The eigenvalues are stored in a diagonal matrix (denoted Λ), where the diagonal elements give the eigenvalues (and all the other values are zeros). Therefore Equation (3-7) can be rewritten as:

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^{-1} \quad (3-10)$$

Since the eigenvectors \mathbf{U} are orthogonal, the matrix \mathbf{U} is an orthogonal matrix [ref.3-16]. This implies the following equality:

$$\mathbf{U}^{-1} = \mathbf{U}^T \quad (3-11)$$

Therefore when we substitute Equation (3-11) for Equation (3-10), it becomes

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^T \quad (3-12)$$

From Equation (3-12) it can be seen that together the eigenvectors and eigenvalues of a matrix constitute the eigen-decomposition of this matrix.

3.6 Computing PCA

The following is a detailed description of how to obtain the PCA if a data set is given.

Step 1: Data set

Organize a data set as a single matrix \mathbf{X} of dimensions $m \times n$. The data set comprises an observation of m variables and n dimensions, or n observations each with m variables.

Step 2: Subtract the mean

The mean subtraction is performed along each dimension. This produces a data set whose mean is zero.

- Calculate mean value of each dimension and store in matrix \mathbf{Z} of dimension $1 \times n$.

$$\mathbf{Z}(n) = \frac{1}{M} \sum_{m=1}^M \mathbf{X}(m, n) \quad (3-13)$$

- Subtract the mean matrix \mathbf{Z} from each column of data matrix \mathbf{X} and then store in the $m \times n$ matrix \mathbf{B} .

$$\mathbf{B} = \mathbf{X} - \mathbf{kZ} \quad (3-14)$$

Where \mathbf{k} is $m \times 1$ column matrix of all 1s.

Step 3: Calculate the covariance matrix \mathbf{C} of \mathbf{B} from step (2). This is done

in the same way as described in section 3.5.1. This yields the covariance matrix \mathbf{C} of $n \times n$ dimensions.

Step 4: Compute the eigenvectors \mathbf{U} and eigenvalues Λ of the covariance matrix \mathbf{C} .

- The n eigenvectors of \mathbf{C} along with their eigenvalues.
- The eigenvectors have to be normalised to unit length such that $\|\mathbf{U}\| = 1$.

Step 5: Rearrange the eigenvectors and eigenvalues

- Rearrange the eigenvectors \mathbf{U} and eigenvalues Λ in order of decreasing eigenvalue, maintain the correct pairings between the eigenvectors and eigenvalues.

$$\mathbf{V} = (eig_1 \ eig_2 \ eig_3 \dots eig_n) \quad (3-15)$$

Where eig_i is i th eigenvector ordered from highest to lowest eigenvalue.

- This gives the components in order of significance, the eigenvector with the highest eigenvalue is the principal component of the data set.

Step 6: Project the data set onto the new basis

$$\mathbf{Y} = \mathbf{V}^T \mathbf{B}^T$$

Where \mathbf{V} is the chosen eigenvectors from step (5) and \mathbf{B} is the mean subtracted data from step (2).

- The \mathbf{Y} matrix is the final data set with the items in columns and dimensions along rows. Now we have changed the original data from being in term of normal axes, such as x and y , into vectors axes. This will give the original data in term of the eigenvectors. These eigenvectors are always perpendicular to each other.

When the data set have been transformed, the values of the data points tell us exactly where (i.e. above/below, distance) the trend lines the data point stands. In the case of the transformation using p eigenvectors, we have simply altered the data so that it is in terms of those eigenvectors instead of the usual axes.

If we choose only the first p eigenvectors according to the highest p eigenvalues in step (5), then the final data set has only p dimensions. This means that the original data set can be compressed and reduced in dimension.

3.7 An Example of PCA

This example demonstrates the processes of PCA as explained above through the made up data, original data. The data comprises two sets of data, D1 and D2, which are shown in Table 3.1 along with their mean subtracted.

Table 3.1 The original data and its mean subtracted

| Original data | | Mean subtracted data | |
|---------------|-------|----------------------|------------|
| D1 | D2 | $\bar{D}1$ | $\bar{D}2$ |
| 8.2 | 8.8 | 2.57 | 3.18 |
| 6.4 | 7.1 | 0.77 | 1.48 |
| 8.1 | 9 | 2.47 | 3.38 |
| 6.6 | 5.6 | 0.97 | -0.02 |
| 3.4 | 3.6 | -2.22 | -2.02 |
| 2.8 | 3 | -2.82 | -2.62 |
| 3.4 | 2.3 | -2.22 | -3.32 |
| 5.3 | 4.5 | -0.32 | -1.12 |
| 7.2 | 8.2 | 1.57 | 2.58 |
| 3 | 2.8 | -2.62 | -2.82 |
| 8.3 | 8.6 | 2.67 | 2.98 |
| 5.6 | 5.6 | -0.02 | -0.02 |
| 3.7 | 3 | -1.92 | -2.62 |
| 7 | 6.4 | 1.37 | 0.78 |
| 5.4 | 5.8 | -0.22 | 0.18 |
| Mean | 5.626 | 5.62 | |

Step 1:

The matrix \mathbf{X} , $m \times n$ dimensions with $m=15$ and $n=2$, can be formed from the original data as:

$$\mathbf{X} = [D1 \ D2]$$

Step 2:

The mean subtracted data for each dimension can be calculated as explained above in step (2). The mean subtracted data are shown in Table 3.1. The matrix **B** can be formed:

$$\mathbf{B} = \begin{bmatrix} \bar{D}_1 & \bar{D}_2 \end{bmatrix}$$

Step 3:

Since the mean subtracted matrix **B** has two dimensions, the covariance matrix will be 2x2 matrix. This is done as explained in section 3.5.1 and its result is

$$\mathbf{C} = \begin{pmatrix} 3.8907 & 4.4887 \\ 4.4887 & 5.5674 \end{pmatrix}.$$

Step 4 and 5:

From the covariance matrix **C**, the eigenvalues and eigenvectors can be calculated. Then rearrange the eigenvectors and eigenvalues in order of decreasing eigenvalue to form eigenvector **V**. The first column of eigenvector **V** corresponds to the largest eigenvalue Λ , and so on. The first vector, first column, and second vector in **V** are the first and second principal component of the data set, respectively. The graph of mean subtracted data overlaid with eigenvectors **V** is shown in Figure 3.8. This provides information relating to the pattern in the data. The eigenvectors show how

these two data sets are related along that line. The data points follow the main line, PC1, but are off the side of the line by some amount.

$$\Lambda = \begin{pmatrix} 9.2954 & 0 \\ 0 & 0.1627 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} 0.6389 & -0.7692 \\ 0.7692 & 0.6389 \end{pmatrix}$$

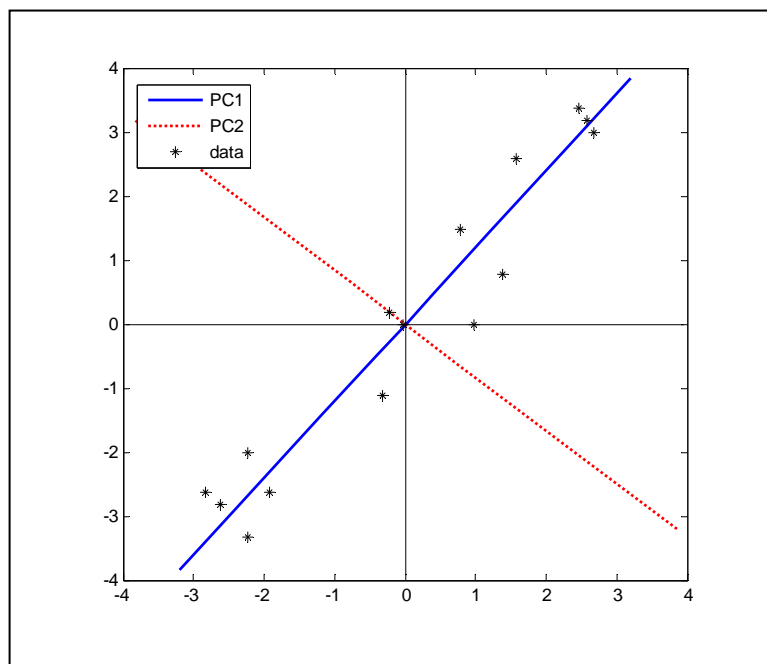


Figure 3.8 A graph of mean subtracted data with the eigenvectors overlaid on top: first principal component (PC1) and second component (PC2)

Step 6:

In this step the new data set \mathbf{Y} is formed and its result is shown in Table 3.2.

This new data was transformed into the eigenvectors axes. The graph of the

new data shows in Figure 3.9. This graph is basically the original data rotated and it is plotted by using the eigenvectors as axes.

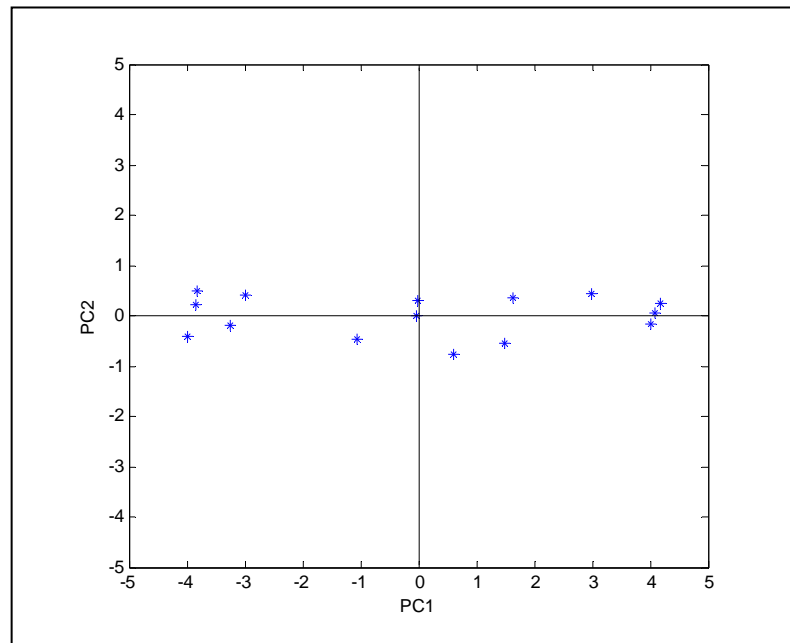


Figure 3.9 Graph shows the new data

Table 3.2: The new data obtained from applying the PCA using all eigenvectors

| Y | |
|----------|--------|
| PC1 | PC2 |
| 4.090 | 0.052 |
| 1.632 | 0.350 |
| 4.180 | 0.256 |
| 0.606 | -0.761 |
| -2.976 | 0.422 |
| -3.821 | 0.500 |
| -3.976 | -0.408 |
| -1.070 | -0.464 |
| 2.989 | 0.438 |
| -3.847 | 0.218 |
| 4.000 | -0.152 |
| -0.032 | 0.007 |
| -3.246 | -0.191 |
| 1.477 | -0.558 |
| -0.006 | 0.289 |

When the data set have been transformed, the values of the data points tell us exactly where (i.e. above/below, distance) the trend lines the data point stands. This is clearly shown in Figure 3.9.

If we choose only the first p eigenvectors according to the highest p eigenvalues in step (5), then the transformed data set has only p dimensions. This means that the original data set can be compressed and reduced its dimensions. In our example, if we choose only the first eigenvector, we will simply have the new data set \mathbf{Y} of one dimension, PC1, as shown in Table 3.2.

Chapter 4

4 Application of Hough Transform and PCA

The basic operations of both the Hough transform and the PCA have been described in the last Chapter. As discussed, they were able to perform the functions as originally designed. However, in order to be applicable to the tasks required for this project, certain modifications are required. In this Chapter, the modification and the application of Hough Transform to our project will be explained first. The details of the modified algorithm for detection of line segments in grey scale images directly will be discussed. The capability of the modified algorithm will be demonstrated through simulations. This will be followed by detailed description of the subsequent application of principal component analysis to our project.

4.1 Lines detection in grey scale image using Hough Transform

The Standard Hough Transform is originally used to detect lines in binary images (binary edge images). For a grey scale image, therefore, it must be pre-processed using an edge detection technique. In our project, we modify the original Hough transform and apply the new algorithm to detect the positions and orientations of line segments in grey scale images directly. In our modified algorithm, we also can measure the line widths and any varying intensities along the lines. These line segments may be in the form of isolated segments, or multiple segments not in

contact with each other, or multiple segments that intersect each other as have been shown in Figure 3.1.

The use of Hough transform to detect line(s) in grey scale images will be explained with the procedures below.

1. Create the **two** parameters (ρ, θ) space as two-dimensional matrix.

These matrices are known as the Accumulator Array 1 & 2 (*Acc1* and *Acc2*). In the original transform, only one is used. The quantisation levels along the (ρ, θ) dimensions are user specified.

2. Remove background by thresholding and possibly other constraints if necessary.
3. Initial the matrices *Acc1* and *Acc2* to zero.
4. Mapping of points in image space to the parameter space (θ, ρ) , as described in Eq. 3-3, and storage in Accumulators. Of the two accumulators, one performs the same function as the accumulator described in Chapter 3, and is therefore used to represent the number of points along the length and width of any line segment. The other accumulator, however, is used to store information relating to the magnitudes, or intensities, of the image points. The elements in the matrices will have their values updated:

4.1 Accumulator 1 : store magnitudes of image points,

$$Acc1 = \sum_{j=1}^n I_j \quad ; \text{ adding magnitude of image}$$

where I_j is the magnitude at pixel j .

4.2 Accumulator 2 : store number of points

$$Acc2 = \sum_{j=1}^n 1 \quad ; \text{ adding number of points}$$

5. Repeating step 4 for all the image points would result in the vote matrices, showing the frequencies of the image points for various (ρ, θ) values.
6. Interpretation of accumulators $Acc1$ and $Acc2$ to yield the line segments. The interpretation is done by thresholding and possibly other constraints where only areas with votes greater than a certain value are taken. These areas correspond to lines in the original image and the value of the area provides a measure of the number of points, or the length of the line, and the width of line.
7. Conversion of each peak area to a thin line (option).

4.2 Transformation to Hough space

The Hough transform takes a grey scale image as input. Every image pixel is transformed to all possible lines that could pass through that pixel. Figure 4.1 illustrates this for a single line image. When a pixel is transformed, bins in both accumulators are increased for all lines that could pass through that pixel. Consider two points **P1** and **P2** on the line segment with magnitudes of 1 and 0.5 respectively. **P1** will produce a sine wave **C1** according to Eq. 3-3, and **P2** another sine wave **C2**. These two sine waves are shown in Figure 4.1 (b). For accumulator $Acc1$, **C1** has

the value of 1 and **C2** is 0.5. The cumulated value at the intersection **J** is therefore 1.5 plus any previous value. This means that the increase in the bins of the accumulator *Acc1* each time is the value of each sine wave at the intersection point. In the accumulator *Acc2*, both sine waves **C1** and **C2** will have values of 1, thus the cumulated value at the intersection **J** will be 2.

Figure 4.3 (a) is obtained by taking the transforms of all the image pixels in figure 4.1 (a). In this case, the brightness of each sine wave is made proportion to the intensity of the pixel concerned. Thus figure 4.3 (a) corresponds to the accumulator *Acc1*. Figure 4.3 (e) is also resulted from the transformation of the original image in figure 4.1 (a), but with the brightness of the sine wave having value of 1. This results in the accumulator *Acc2* to be generated.

Now that the accumulations of *Acc1* and *Acc2* are completed, they are used to determine the characteristics of the original line segment. As have been mentioned before, the two accumulators *Acc1* and *Acc2* serve different functions. *Acc1* is used to determine the location, or (ρ, θ) of the line segment, whereas *Acc2* is used to measure the width and length of line segment. Detailed applications of the two accumulators will be discussed in sections 4.3 and 4.4. The remaining of this section will be devoted to discussing the characteristics of the outputs of the two accumulators.

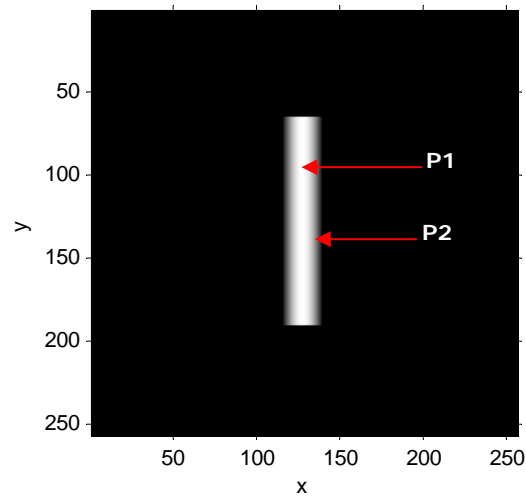


Figure 4.1 (a) Line image : the angle is 0° (vertical line)

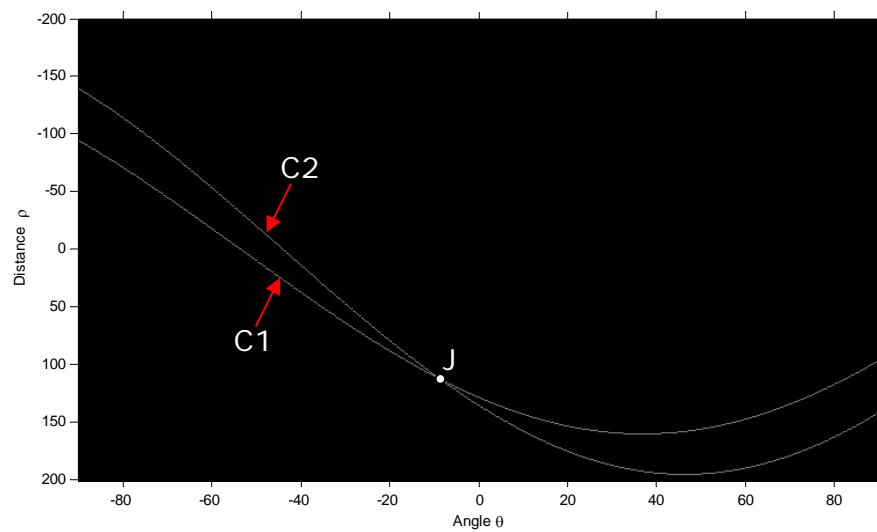


Figure 4.1 (b) Hough space: Two sine waves are transformed from points
P1 and **P2**

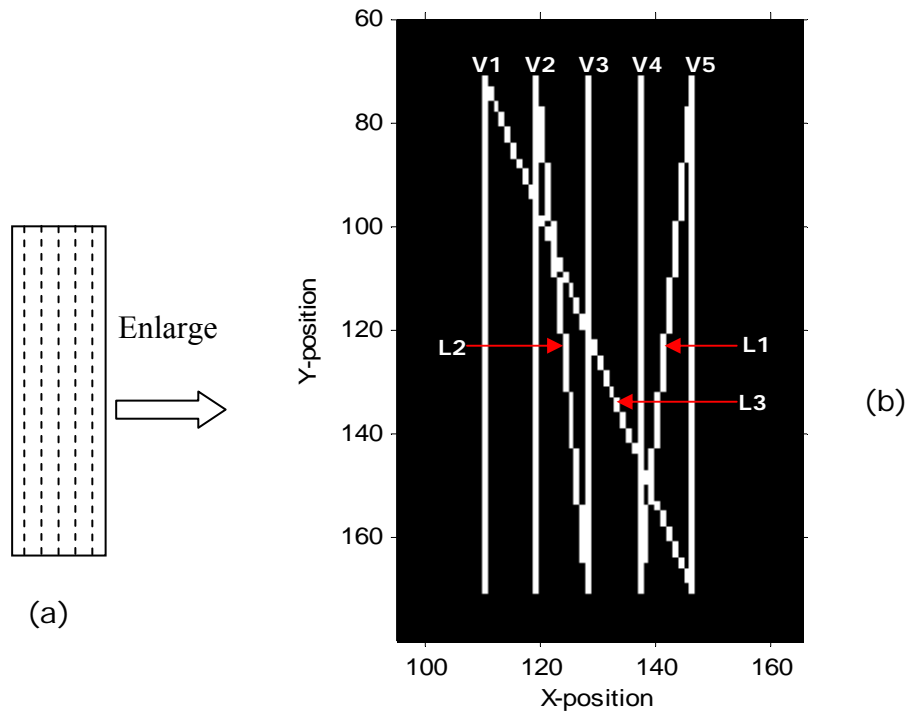


Figure 4.2 (a) A small line segment, (b) Enlarged thin line segments of (a) (one pixel wide)

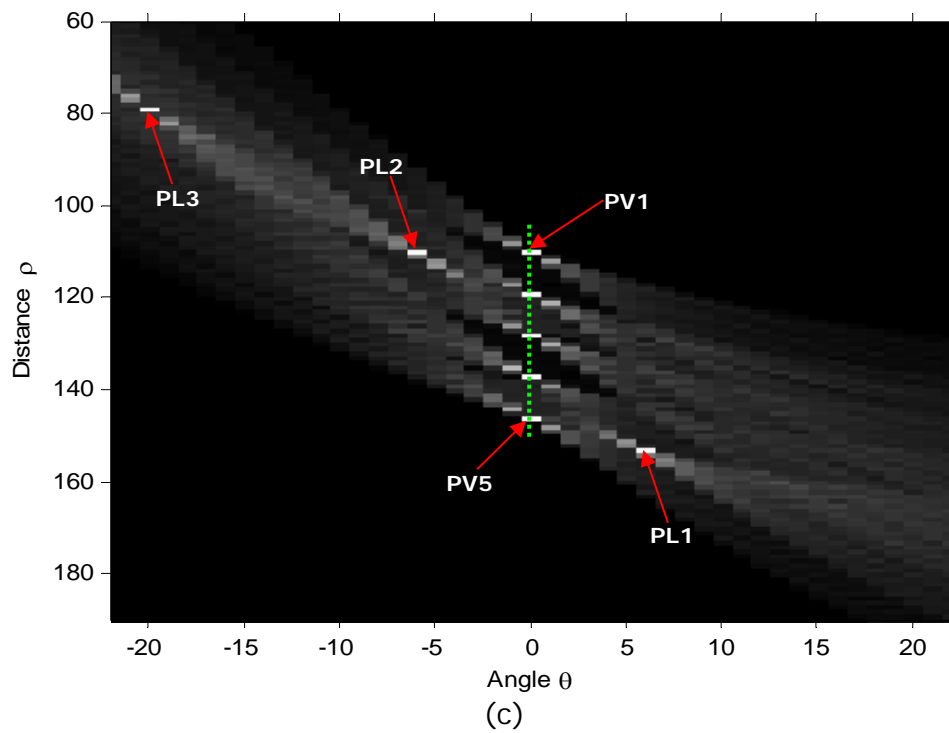


Figure 4.2 (c) Magnified Hough space obtains from transforming of (b)

Now consider figure 4.2 (a) which shows a small line segment. In Matlab, this is represented by an array of values. Therefore, at the enlarged figure (b), each vertical line segment will correspond to a dashed line in (a), or one column of data in Matlab. As shown, there are five vertical lines from **V1** to **V5**. Considering line **V1** first, it has an angle of zero, and a certain distance ρ from the origin. It will therefore be transformed to **PV1** at the transform space. Similarly, the other four lines **V2** to **V5** will result in **PV2** to **PV5**. Together they form a straight line (in green) in the transform space at $\theta = 0^\circ$. On the other hand Hough transform will consider any two points within the matrix capable for forming a straight line. For example, **L1** in (b), which is formed by the two end points of **V4** and **V5**, will result in **PL1** in the transform space. **PL2** and **PL3** thus formed similarly. Indeed, any combinations of points in (b) will feature in the transform space. The crucial point is that, only the **V** lines will form features in the transform space along the 0° angle direction (or other directions according to the original line segment), features due to the **L** lines will spread around the green line in the transform space, and the cumulated values for the two accumulators will have their maximum on the green line. It is this aspect of the Hough transform that we exploit to determine the orientation and location of lines in the original image.

4.2.1 Accumulator 1 : ACC1

Consider the brightest area (region that is of the highest cumulated values) of accumulator *Acc1*, which is indicated inside the rectangular box in figure 4.3 (a), and a magnified version is shown in figure 4.3 (b). Figure 4.3 (c) is derived from (b). Considering the curve in black first, it represents the profile of the *Acc1* output at angle = 0° . In other words, it is the cumulated values along the green line in figure

4.3 (b). The other curves in (c) are similarly extracted from (b), but for angles between -6° to 6° . The graph therefore plots the magnitudes of *Acc1* output as a function of the distances of the line segment. It should be noted that the original line segment in figure 4.1 (a) has an angle of 0° , and this angle is translated directly to figure 4.3 (b & c). Therefore, by determining the curve in (c) that contains the maximum value, we have determined the angle of the original line segment. The distance ρ of the line segment can now be determined which is simply the location of the peak along the black curve. In this example, the maximum magnitude is 125 and this occurs at distance ρ of 127.

Figure 4.3 (d) also shows the profiles of accumulators *Acc1* at the same area as in figure 4.3 (c), but these profiles are taken from the angle 0° to 10° and shifted to middle. The graph shows clearly that the magnitude of each profile gradually decreases when the angle of the selected profiles deviates from the correct angle (0°). This further demonstrates that the maximum point in the brightest area is the location of a line segment.

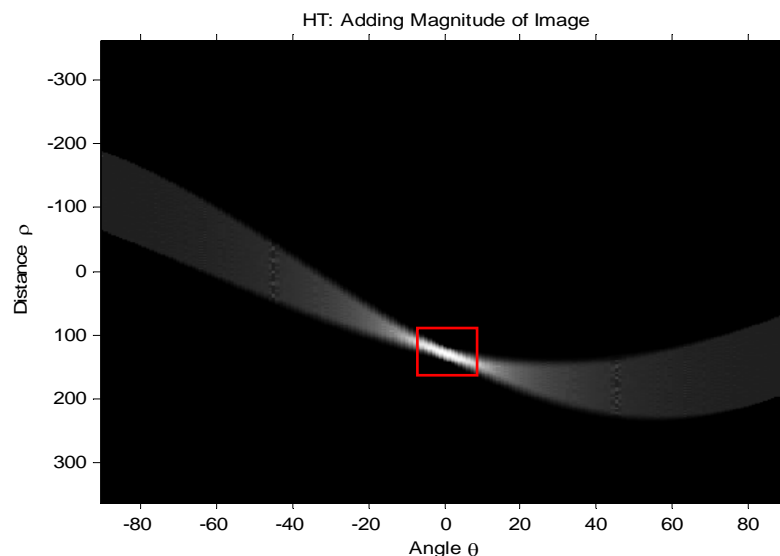


Figure 4.3 (a) The entire Hough space: Adding magnitude of image (*Acc1*)

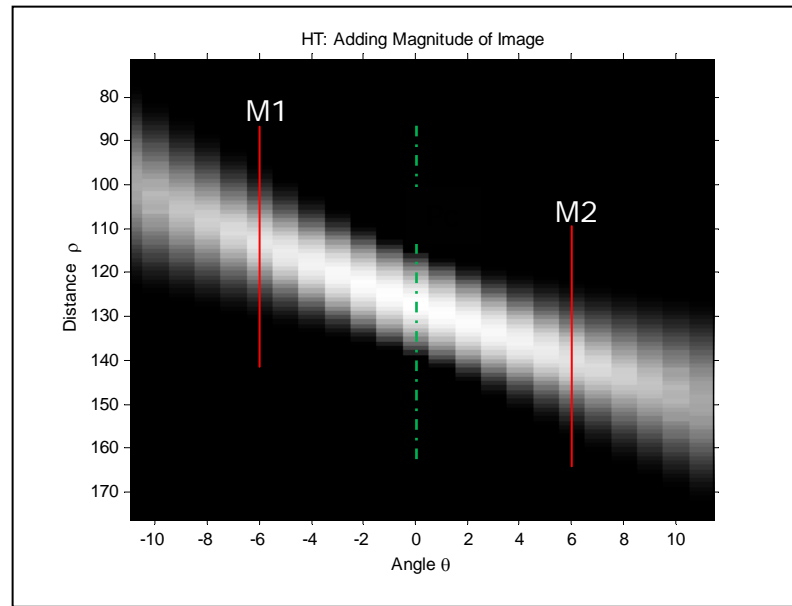


Figure 4.3 (b) The area zooms in on rectangular red box from (a)

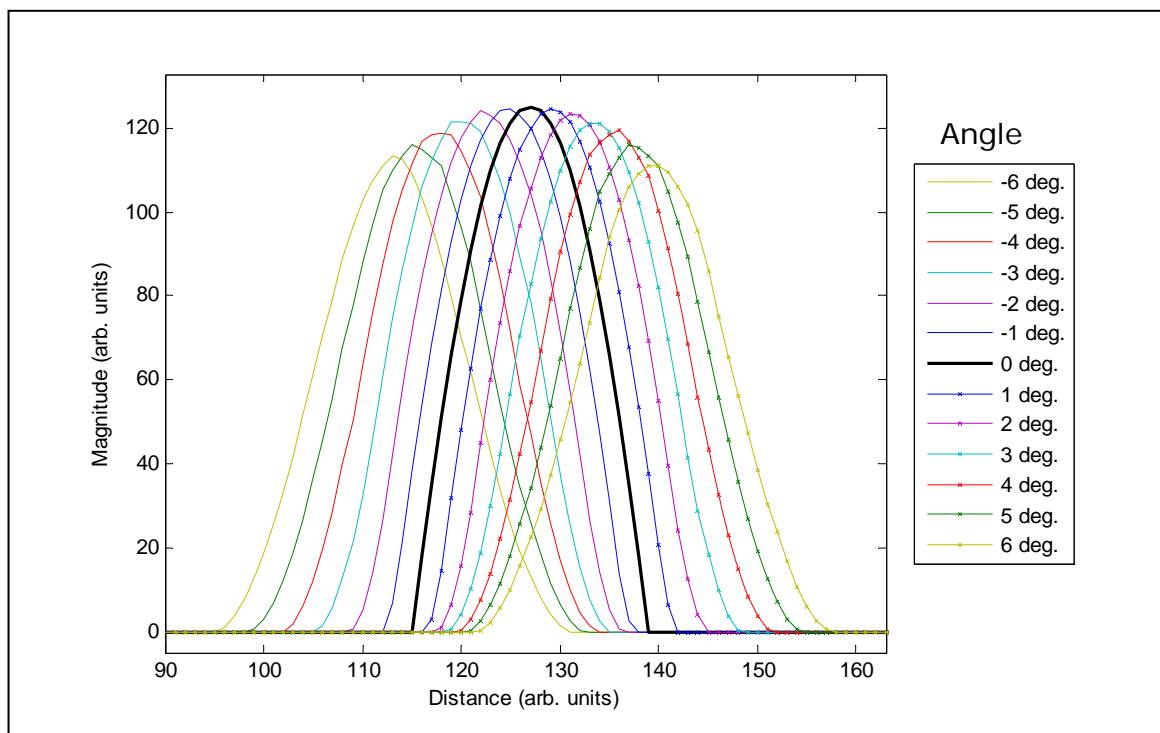


Figure 4.3 (c) The profiles taken from the Hough space between **M1** and **M2** in (b)

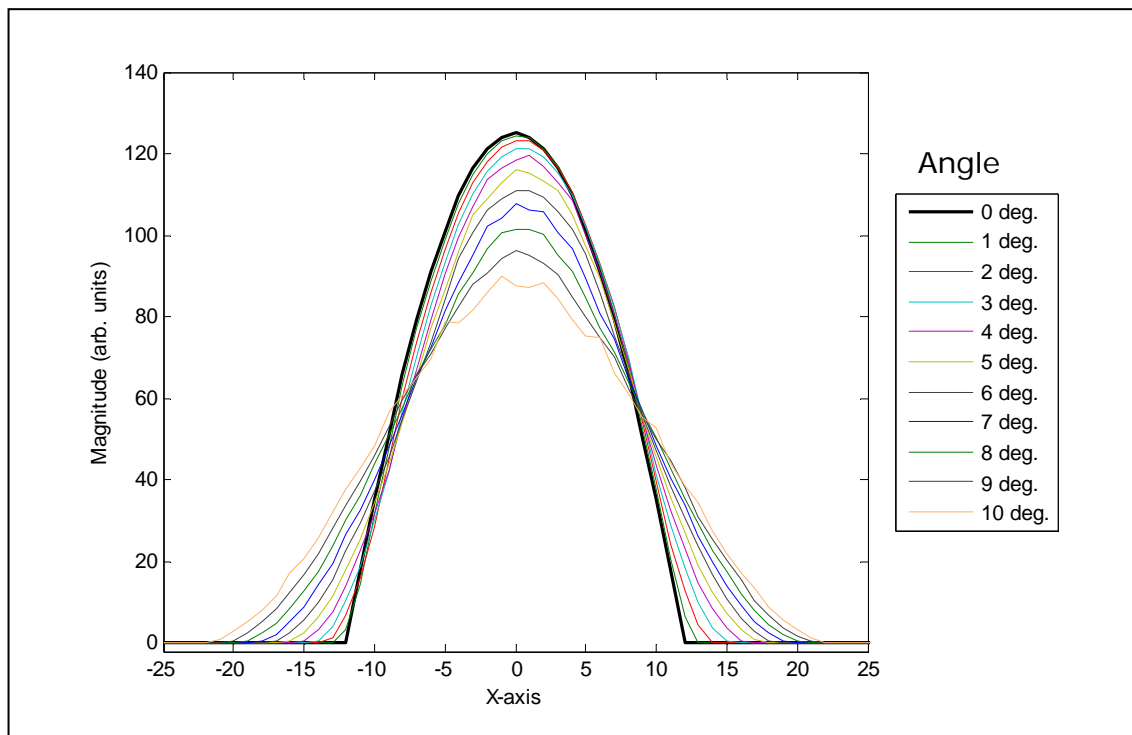


Figure 4.3 (d) Comparing profiles for each angle (0-10 deg.) of *Acc1*

4.2.2 Accumulator 2 : ACC2

The function of *Acc2* is to produce information relating to the width and length of the line segment. Consider the brightest area of accumulator *Acc2*. Figures 4.3 (e - h) are the equivalents of (a - d) for *Acc1*. The difference between the two cases is illustrated in (g) and (c). From figure 4.3 (g), it can be seen that the top of each profile is flat and the maximum magnitude is approximately the same. However, the widths at the top of these profiles are different. Figure 4.3 (h) is the evidence of this. The graph clearly shows that the width at the top of each profile gradually decreases when the angle shifts away from its actual angle (0°). This means that the profile at the actual angle will be wider than others. Thus the width of this profile considered to be the width of line segment.

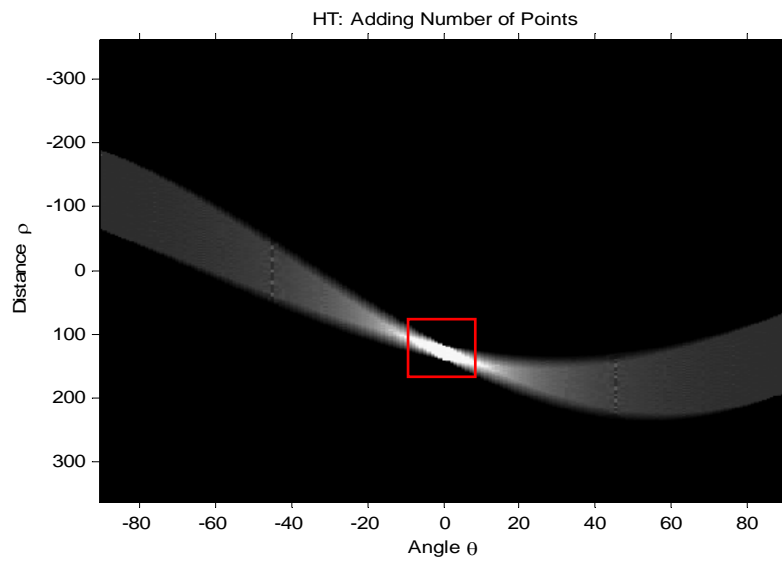


Figure 4.3 (e) The entire Hough space: Adding number of points (*Acc2*)

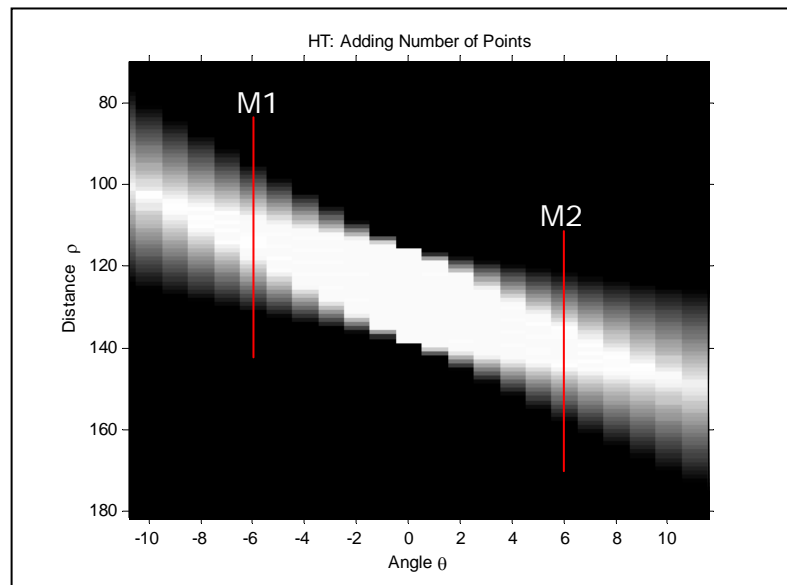


Figure 4.3 (f) The area zooms in on rectangular red box from (e)

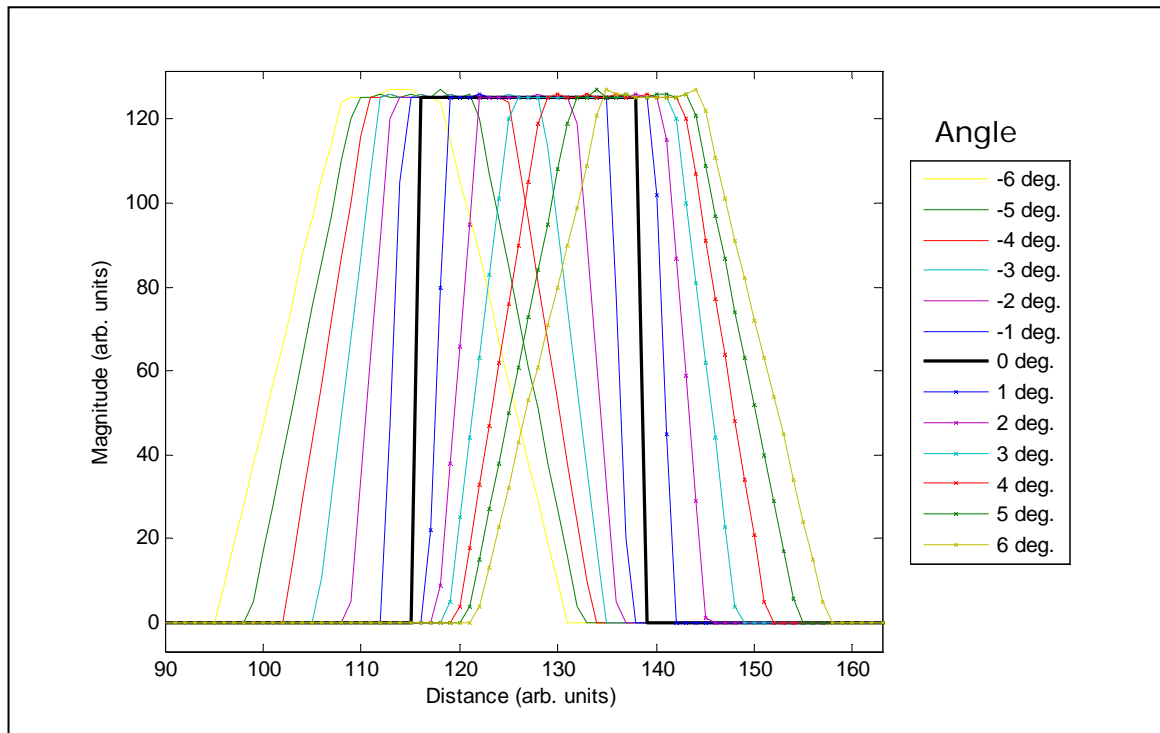


Figure 4.3 (g) The profiles taken from the Hough space between **M1** and **M2** in (f)

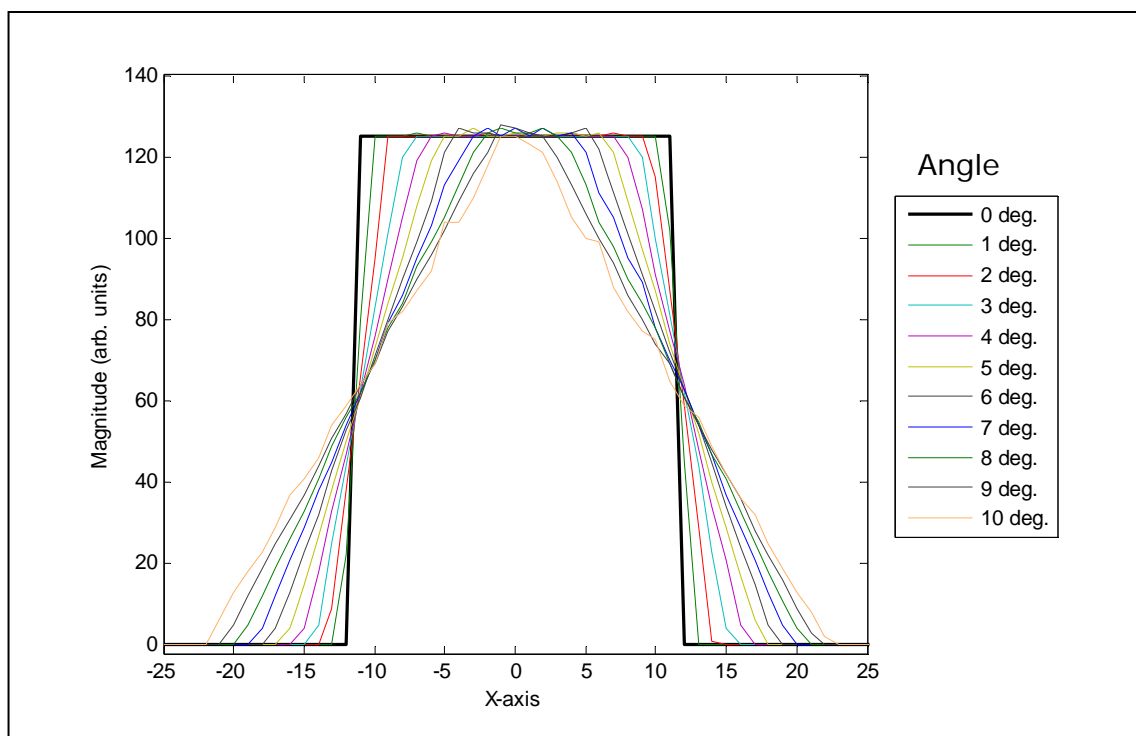


Figure 4.3 (h) Comparing profiles for each angle (0-10 deg.) of *Acc2*

The following example demonstrates relationship between the width of original line segments in figure 4.4 (a) and its corresponding transform in accumulator *Acc2*. The solid curves in figure 4.4 (b) are derived from (a) and the dash-dot curves are obtained from the output accumulator *Acc2*. Considering both curves in blue first, the solid one represents the profile of **L1** along the red line. Obviously, it is the intensities or signals strength across the width of the line segment. The blue dash-dot curve denotes profile of the transform of **L1** in accumulator *Acc2* taking at the correct angle, 0° . The other curves in (b) are similarly extracted from (a) and their transform, but for **L2** to **L4** from inner to outer. It can be seen that the width at the base of each solid curve and the one of dash-dot curve, with the same colour, are exactly equal. Therefore, by determining the width of dash-dot curve in (b) at the correct angle, we have determined the width of the original line segment.

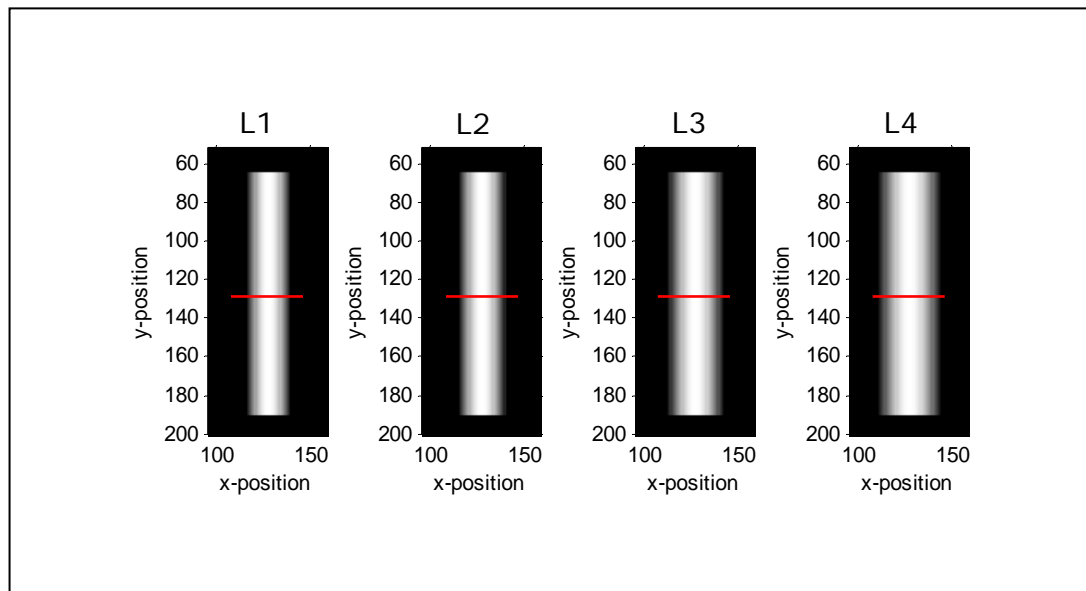


Figure 4.4 (a) original line segments

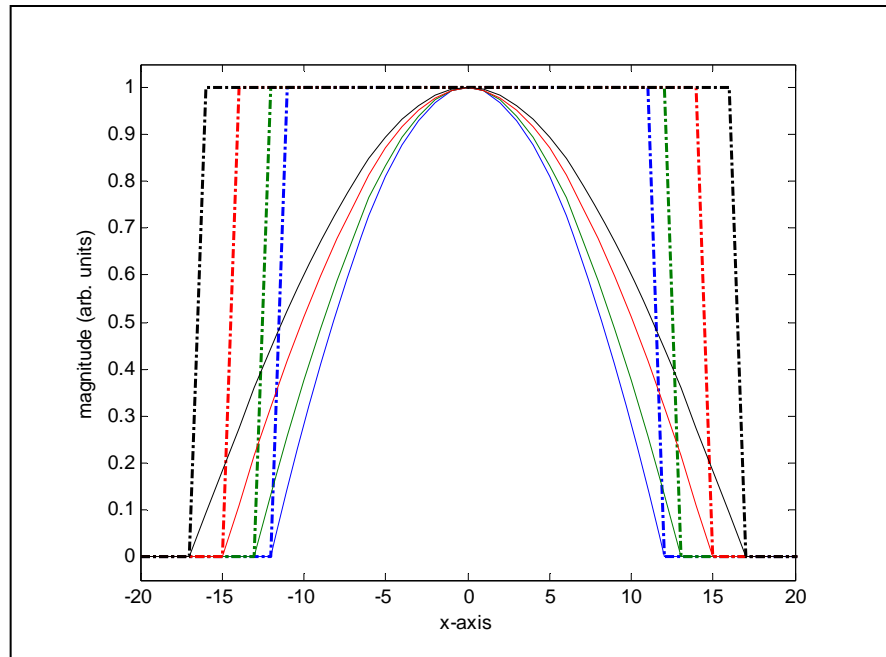


Figure 4.4 (b) profiles of original line segments (solid lines) and of accumulator *Acc2* (dash-dot lines)

4.3 Location of Lines in Accumulator

We have illustrated the procedure for determining the presence of isolated single line in the last section. The procedure for determining multiple lines is similar except that the areas of localised maxima will need to be separated first. This section will discuss aspects relating to this case. The following example uses two line segments to demonstrate the processes. The original image in Figure 4.5 (a) contains two line segments **L1** and **L2** with different positions and angles. They may be of different widths and lengths as well. The transformation of the image yields the Hough space (parameter space) shown in Figure 4.5 (b). It can be seen that there are two bright areas indicated by **A1** and **A2**, the regions with the most intersected points. To

isolate the two regions, we firstly look for a global maximum. A region surrounding this maximum is formed by applying a threshold in the neighbourhood. A value of 80% is used for the thresholding although other values can be used depending on application. Once this area is determined, it is transferred to 4.5 (c) and the process is repeated for the next maximum. As there are two lines in the original picture, figure 4.5 (c) contains two isolated regions. These two areas are processed separately using the method described in the last section, in order to determine their locations and orientations. For example used, the angles and distances are calculated to be $(15^\circ, 50)$ and $(0^\circ, 128)$ respectively.

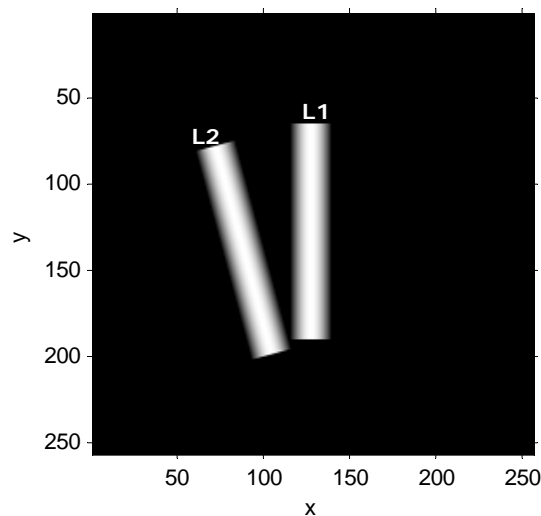


Figure 4.5 (a) Original Image

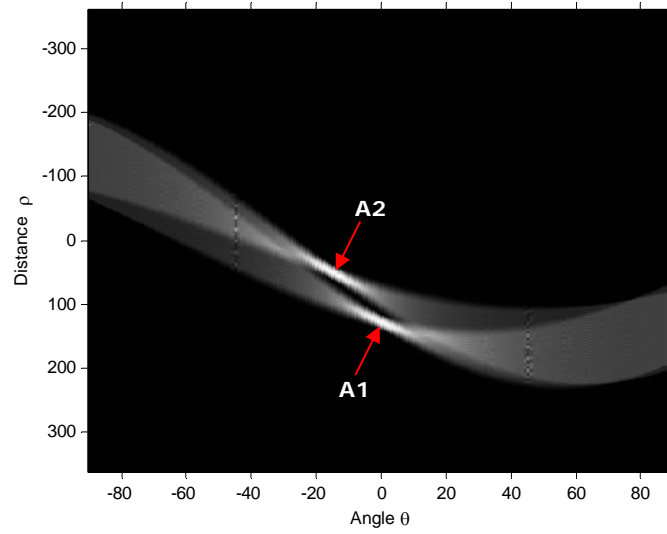


Figure 4.5 (b) Hough space: Adding magnitude of image ($Acc1$)

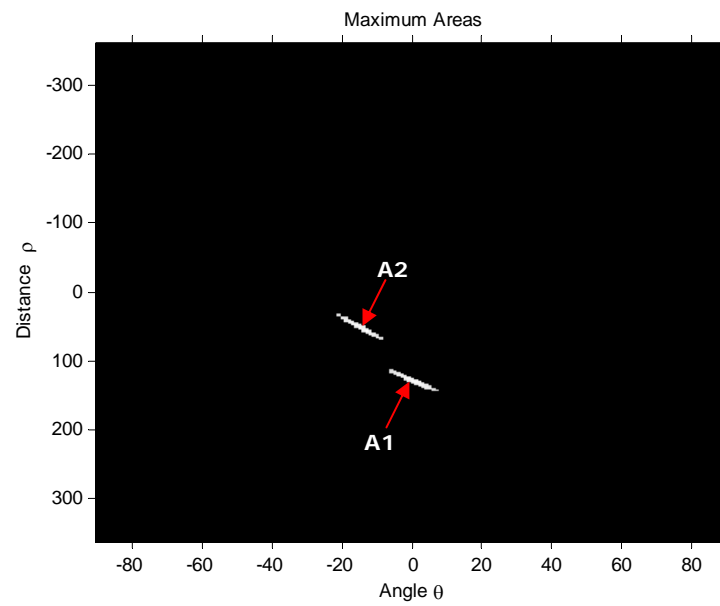


Figure 4.5 (c) Local peak areas

Figure 4.5 The detected lines are shown in the white areas: **A1** and **A2** represent **L1** and **L2**, respectively

Procedure to locate line segments can be summarised as follow:

1. Searching for the global maximum point(s) in accumulator *Acc1* - keep this to be a *maximum reference*.
2. Locate maximum area by applying a threshold in the neighbourhood of global maximum point in (1) – keep only the values greater than 80% of max.
3. Determine distance and angle as have been described in section 4.2.1
4. Set the region in (2) to zeros
5. Repeat step (1)
6. Stop the process if the peak value of remaining area is less than 20% of the *maximum reference*.

4.4 Length and width of lines

The length and width of line segments can be obtained from the accumulator *Acc2* as has been mentioned earlier. Variations in the length of a line affects magnitude of accumulator *Acc2*, whereas the line width changes the width of the bright area at the angle of the line segment, as illustrated in figure 4.6. Figure 4.6(a) shows the accumulator *Acc2* of the original line segments in figure 4.5(a). Once the locations and angles of the two segments are determined using data from *Acc1*, profiles along the peaks of each region can be drawn. Figure 4.6(b) shows the profile of *Acc2* along the peak of region **A1**, at the angle 0^0 . From the profile, it can be determined that the corresponding line segment has a length of 127 pixels and a width of 24 pixels. Applying the same procedure to region **A2** will yield another profile. As the

two original line segments have the same lengths and widths, the profile for **A2** will be identical to that of **A1** but shifted to angle 15° .

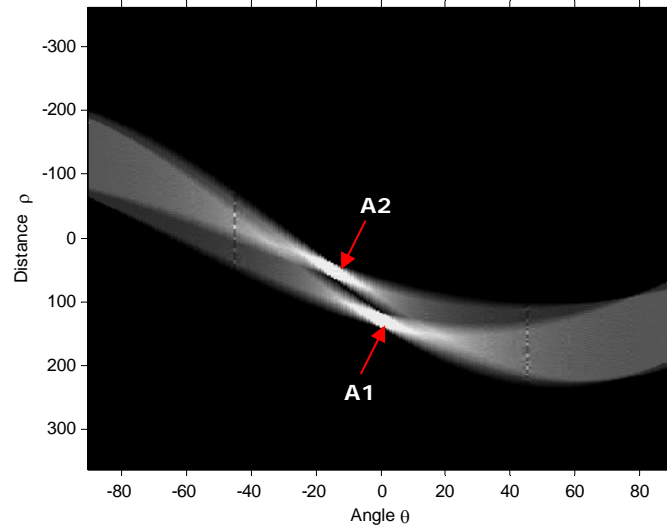


Figure 4.6 (a) Hough space: Adding number of points *Acc2*

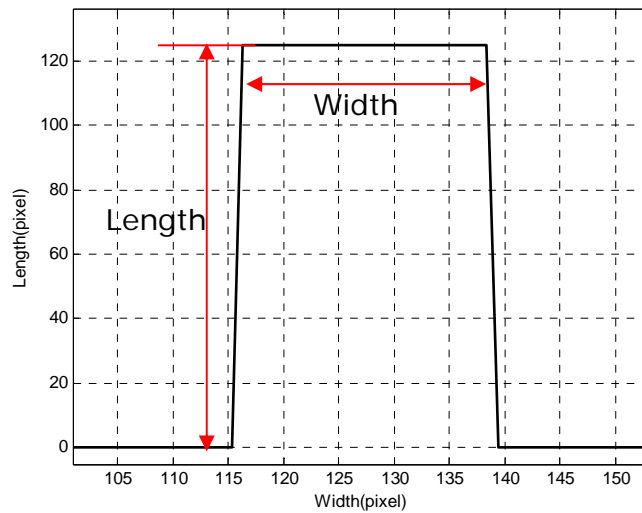


Figure 4.6 (b) Profile at 0°

Figure 4.6 The accumulator *Acc2* in Hough space and its profile at 0° , Original image Figure 4.5 (a)

4.5 Adjacent lines Detection

In some cases the image is formed with two or more line segments situated very close to each other so that the optical system is not capable of resolving them. Figure 4.7 (a) shows one such image formed with two line segments, and (b) is the corresponding profile. Looking at the image, it appears that only one line segment is involved. However, in this section, we will investigate the Hough transform to see if it can provide further information, allowing us to distinguish lines that are closely spaced. In the simulations below, we have included the effects of photon shot noise, which is the dominant noise source with the type of optical imaging system we deal with here.

The simulations are based on the following optical parameters:

- System NA=0.4
- $\lambda = 0.6328 \mu m$
- CCD pixel size = $5 \mu m$
- System magnification = 130

The equivalent pixel width at the sample is therefore is 38.6nm. Figure 4.7 (a) is the image obtained from using a scalar diffraction program, based on the Fresnel approximation, of two closely spaced line segments. After the formation of the image, suitable amount of noise is added to the image. We have only modelled the effects of photon shot noise as its magnitude is much greater than other noise sources

such as dark noise and read-out noise associated with a CCD camera. Different levels of noise, typical of those present in an optical microscope, have been included.

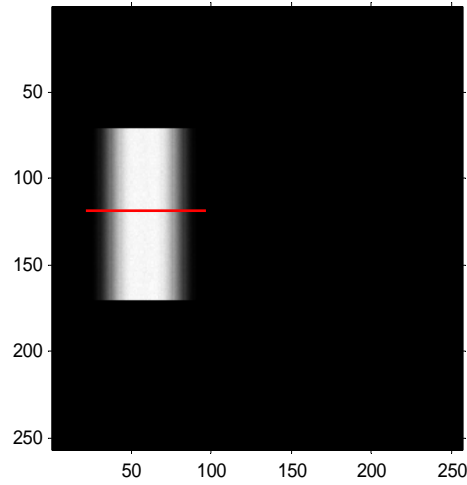


Figure 4.7 (a) Image of two line segments that are very close to each other. Width of picture = $10\mu\text{m}$

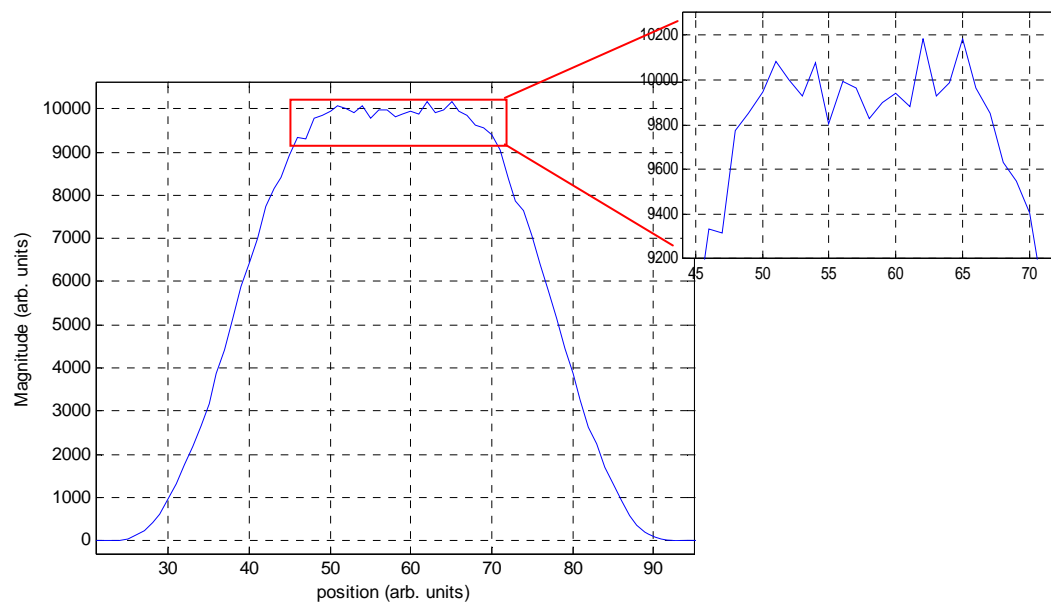


Figure 4.7 (b) profile of (a) along red line

The procedure for examining lines can be divided into the following steps:

- 1) Perform HT on the picture obtained (figure 4.7 (a))
- 2) Locate the maximum point at the peak area in the transformation space
- 3) Take profile from the accumulator *Acc1* at maximum point along angle θ
- 4) Filter the profile from step (2)
- 5) Differentiate the filtered profile obtained from step (3)
- 6) Determine the global maximum and minimum points of the result in step (4)
- 7) Examine the changes of the signs between maximum and minimum points obtained from step (5). If the changes, from positive to negative or vice versa, occur twice that means it has two lines and so on.

The entire process is shown in Figure 4.8. From Figure 4.8 (a), the angle of the maximum point is determined, which is 0° for this case. A profile is then drawn along $\theta = 0^\circ$ as represented by the green line in Figure 4.8 (b), which is shown in Figure 4.8 (c). 4.8 (d) is the magnified version of the peak of the profile. As can be seen there are some ripples on the top of the graph, showing the presence of noise. Depending on the noise level, it will be difficult to determine the number of lines by using this profile directly. Crucially, the top of the graph is used to determine the number of lines. A 1x3 moving average filter is therefore applied to the profile. It results in the smooth traces as shown in Figure 4.8 (e & f). After the filtering process the profile is differentiated, resulting in Figure 4.8 (g). The crucial point in determining the number of lines is the alternation of the sign, positive and/or negative, in graph between the global maximum and minimum points. Consider Figure 4.8 (g) first, from the graph the global maximum and minimum points can be

obtained which are represented by **M1** and **M5** respectively. Then consider Figure 4.8 (g) together with (h), magnified version of (g) in red box, we will have the changes of the signs, M1 to M2 (+), M2 to M3 (-), M3 to M4 (+) and M4 to M5 (-). There are two changes in sign (count in pairs) therefore it has two lines in the area. It should be note that the graph does not oscillate between any two points, for instance, **M2** and **M3**. In the situation when large amount of noise is present in the signal, the identification of multiple lines would fail. The minimum SNR for that to occur will be discussed in the next section.

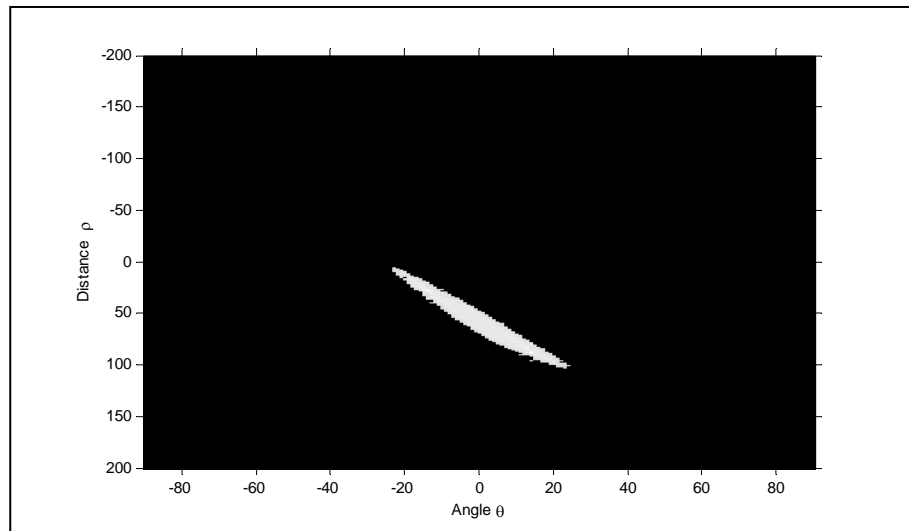


Figure 4.8 (a) The peak area of the HT of a line segment

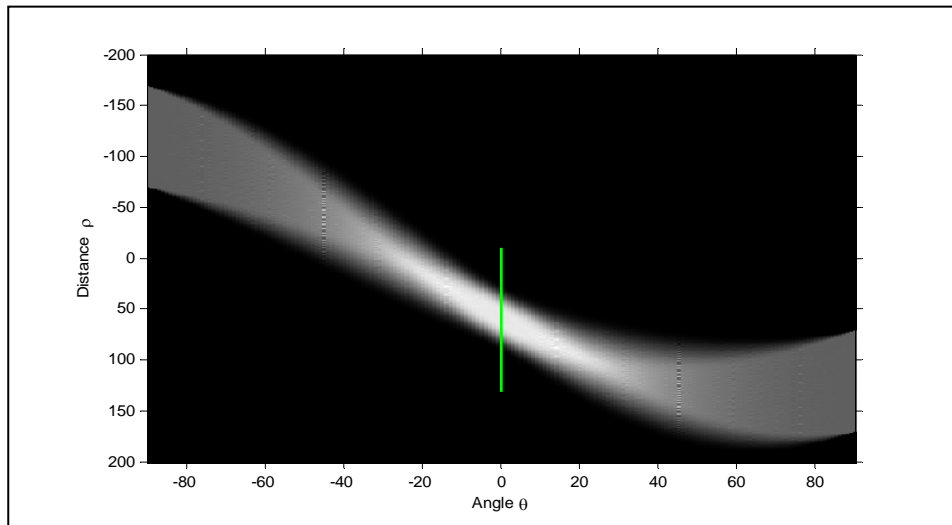


Figure 4.8 (b) The Hough space $Acc1$, original image Figure 4.7 (a)

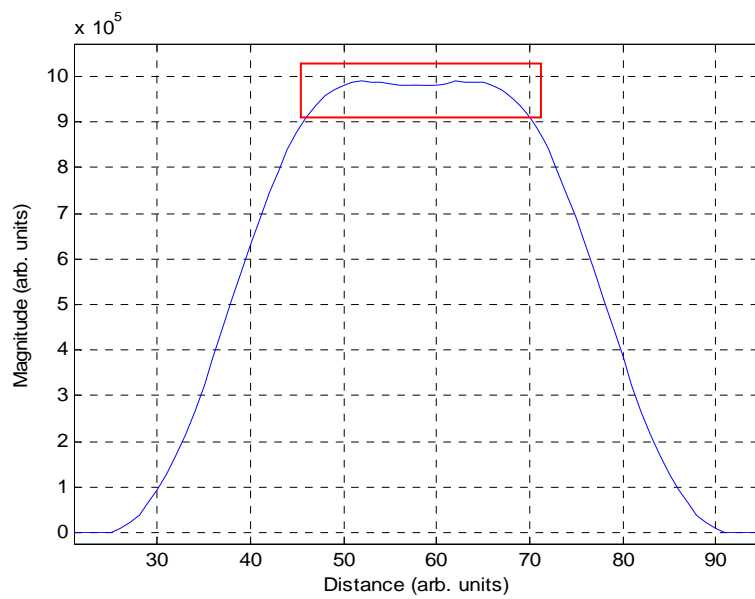


Figure 4.8 (c) Profile takes at 0° , green line in (b)

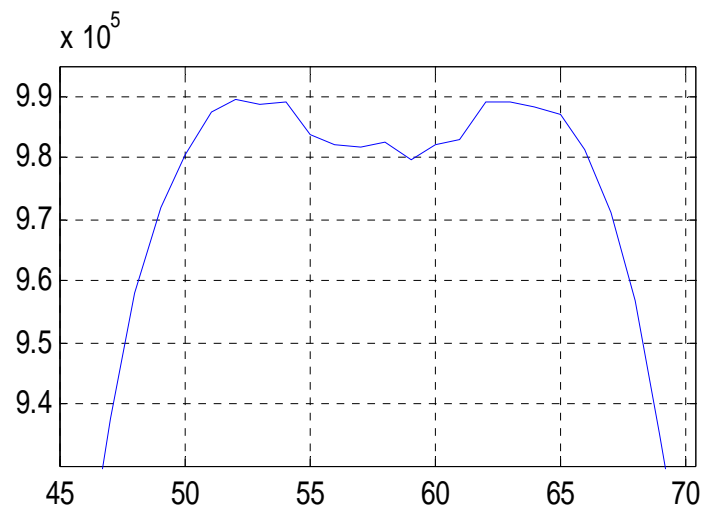


Figure 4.8 (d) Enlarged profile of (c) in red square box

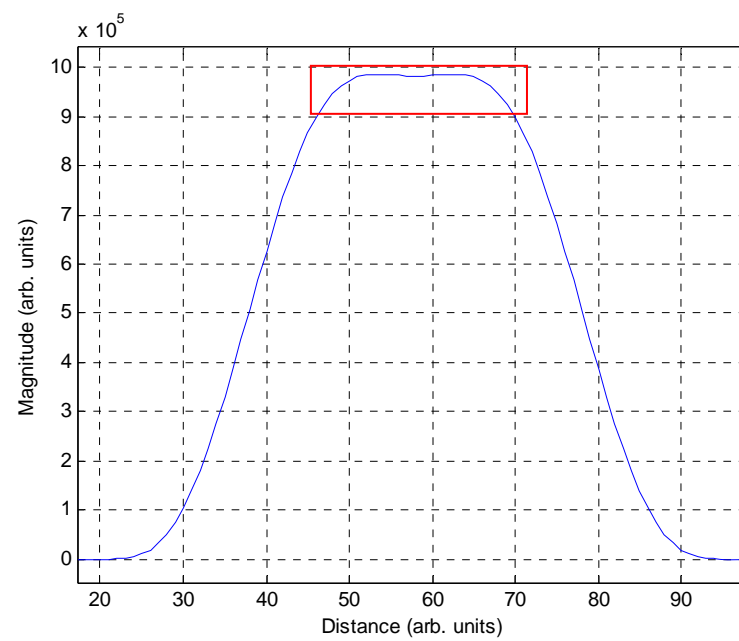


Figure 4.8 (e) Filtered profile of (c)

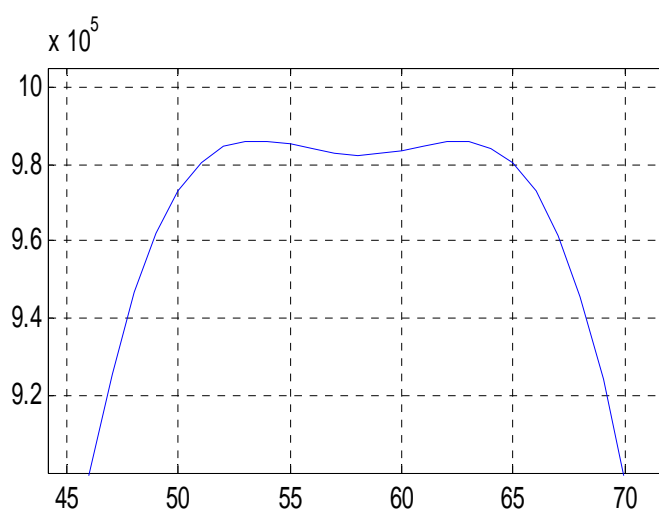


Figure 4.8 (f) Enlarged profile of (e) in red square box

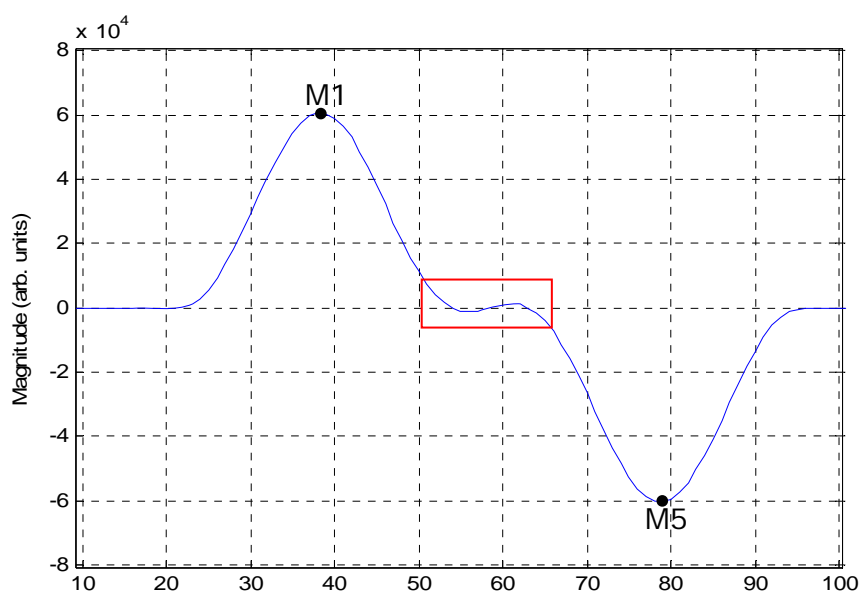


Figure 4.8 (g) Differentiation of filtered profile in (e)

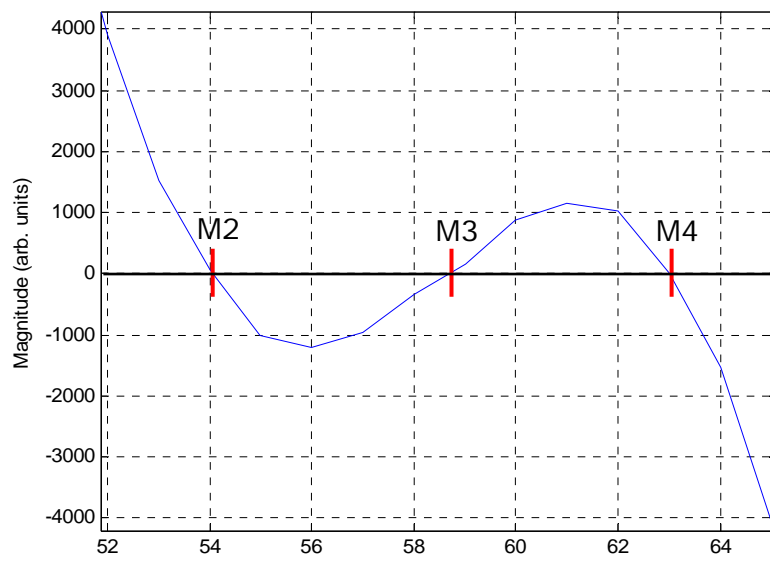


Figure 4.8 (h) Enlarged profile of (g) in red square box

Figure 4.8 Determining if two lines are very close in an image, Original image

Figure 4.7

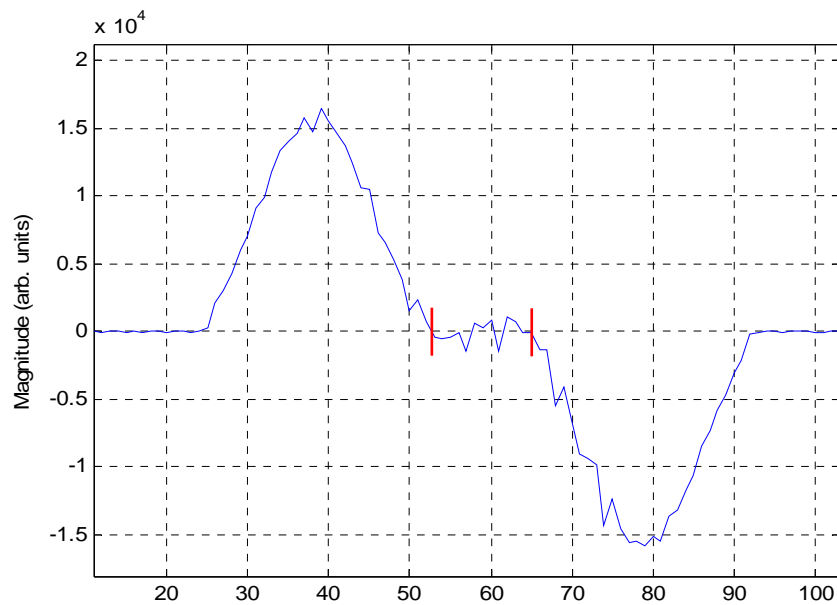


Figure 4.9 Differentiation of filtered profile with SNR=50,

line distances = 19 pixels

From Figure 4.9, the graph derives from the original image, figure 4.7 (a), with signal to noise ratio (SNR) equals 50. It can be seen that there is some fluctuation along the graph thus we cannot determine number of lines in this case. The Table 4.1 shows the performance of the technique in term of signal to noise ratio. The technique can distinguish two lines with the separation down to 19 pixels and the SNR=100.

Table 4.1 performance of the technique

| Line Distances (pixels) | Number of Photons | SNR (x100 averages) | Distinct |
|----------------------------|----------------------|------------------------|----------|
| 19 | 2500 | 50 | × |
| 19 | 6400 | 80 | × |
| 19 | 10000 | 100 | √ |
| 19 | 14400 | 120 | √ |
| 19 | 22500 | 150 | √ |

It should be noted that although we can determine the number of line segments, in case they are very close to each other, the width of individual line segment cannot be measured. The Principal component analysis (PCA) therefore will be used to do this.

4.5.1 System noise sources

Photon shot noise

The level of photon shot noise is calculated in terms of the saturation electron level of the CCD camera. The procedure is simple: the simulated detector output is first

normalised with respect to the pixel of the maximum output, and the normalised output is then scaled up by N , where N is the saturation electron level per pixel of the CCD array. The output is therefore expressed as number of photons (or electrons) each pixel. Photon shot noise is then added to detector output according to equation 4-1.

$$I_n = I + \sqrt{I} \times randn \quad 4-1$$

The Matlab command *randn* generates a set of normally distributed random numbers with zero mean and a standard deviation of one. Multiplying it by \sqrt{I} will change the standard deviation to the square root of the number of photons detected each pixel, which is of course the characteristic of photon shot noise.

Readout noise

Readout noise is the form of electronic noise described in terms of the number of electrons per pixel. The noise is produced during the transfer of charges in a pixelated detector. Due to the imperfection of the charge transfer process, not all the electrons are transferred from one pixel to the next. Some electrons may remain trapped in the pixel, others may recombine with holes. Modern CCDs have charge transfer efficiency in the range of 99.999% [66]. This means that fewer than one out of 100000 electrons will be left behind. It can be seen that this level of noise is much smaller than that of the photon shot noise, and can be ignored in our application.

Dark noise

With a CCD detector, dark noise is expressed in terms of electrons per unit time at a given temperature. In the photoelectric effect electrons generated by the thermal energy in the environment cannot be distinguished from the electrons generated by photons (signal). Therefore, there is some number of electrons stored in the pixel that are not the result of photons go into the detector. These electrons are generated independent to the light level hitting the detector. Typical value of dark noise at room temperature is $10e^-$ over 0.1 ms integration time. The dark noise can be reduced by cooling the detector to a lower temperature [66].

Vibrations

The levels of vibration that affect an optical imaging system depend on environmental condition (for example doors opening/closing) and the movement of its mechanical components during scanning such as objective, camera and sample stage, and bench top design of the system. The vibration can be expressed in the image shift as a percentage of a pixel [65]. The SNR is calculated as the mean value of photons detected for the original point spread function (PSF) divided by the difference in the shifted value and the original value. A vibration level of 0.05% of a pixel gives a SNR of 1 in 500 (the PSF has a maximum value of 50000 photons).

The expected levels of photon shot noise are added to the simulated image to show the capability of the technique. We have only modelled the effects of photon shot noise as its magnitude, maximum number of photons per pixel, is much greater than other noise sources such as dark noise and readout noise associated with a CCD camera. Different levels of noise, typical of those present in an optical microscope, have been included. Table 4.2 shows the signal to noise ratio according to the

maximum number of photons per pixel of CCD camera. For commercial camera, the typical saturation electron level is around 50000 per pixel [59][60] a requirement easily satisfied in Table 4.1.

Table 4.2 Photon shot noise

| Number of Photons/pixel | SNR |
|----------------------------|-----|
| 2500 | 50 |
| 10000 | 100 |
| 40000 | 200 |

4.6 Principal Component Analysis

One of the most important characteristics of PCA as described in chapter 3 is its ability to measure the similarity of two or more sets of data. In the following sections we will describe how we can exploit this capability to meet the requirements of the project.

4.6.1 Measuring the Similarity using PCA

Measuring similarity by using the PCA, we consider the relationship between the eigenvectors and eigenvalues of the data sets. For simplicity we will restrict the consideration to two data sets first, although the argument can be extended readily to multi-dimensions problem. The two data sets are $D1$, a known data set which serves as the template; and $D2$, an unknown data set whose characteristics we want to

measure. Figures 4.10 and 4.11 show examples of $D1$ and $D2$. For illustration purpose, $D2$ is obtained by adding a small amount of noise onto $D1$, and the difference of the two is shown in Figure 4.12. It should be noted that $D1$ is a profile extracted from a line segment image. We will use them initially to illustrate the PCA process, and then how to extract the relevant information for the project.

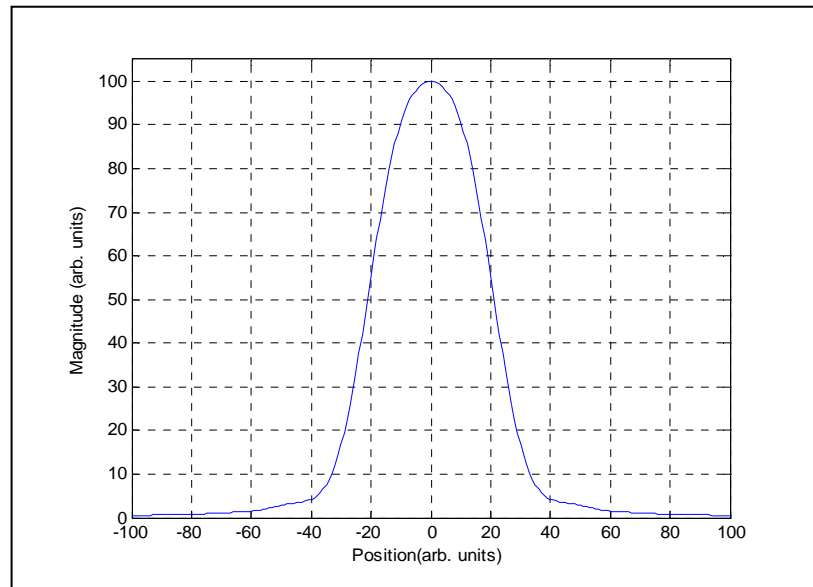


Figure 4.10 Example of $D1$

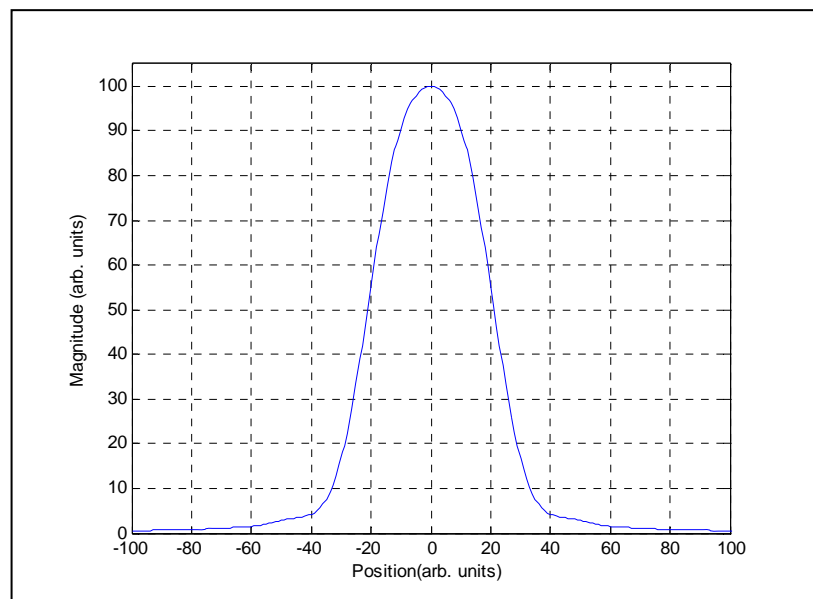


Figure 4.11 Example of $D2$

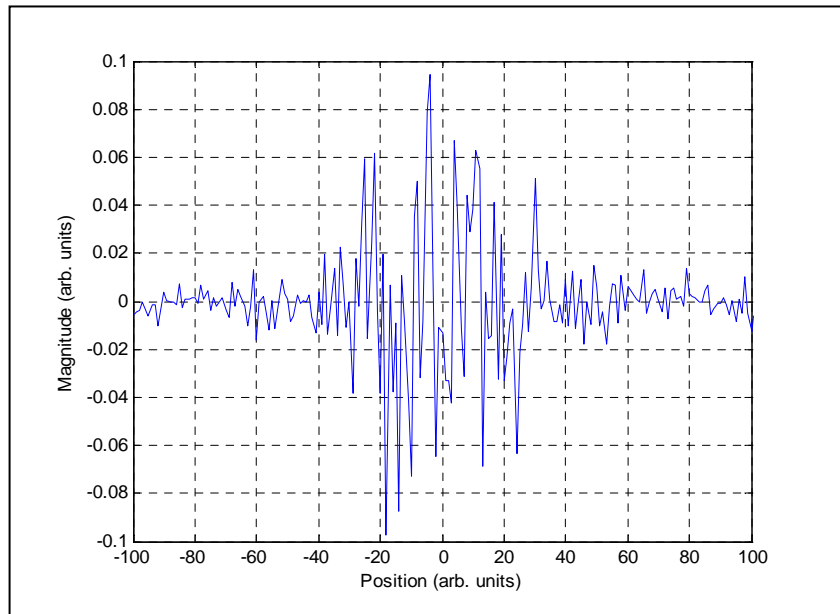


Figure 4.12 The difference of profiles $D1$ and $D2$

The procedures of obtaining the eigenvalues and eigenvectors of the system have been described in Chapter 3. Briefly to remind ourselves

1. Calculate mean subtracted for each data

$$\bar{D}1 = D1 - \text{mean}(D1)$$

$$\bar{D}2 = D2 - \text{mean}(D2)$$

2. Calculate the covariance matrix C of $\bar{D}1$ and $\bar{D}2$
3. Determine eigenvector and eigenvalue from C

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} ; \text{Eigenvector}$$

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} ; \text{ Eigenvalue.}$$

Eigenvalues are the diagonal elements of matrix Λ . The Λ_{11} and Λ_{22} are the minimum and maximum eigenvalues respectively. The eigenvalues show the degree of correlation between the two sets of data. However, their actual values depend on factors such as the number of data points involved, and it is often a better indication by taking the ratio of the two eigenvalues. The calculation of the eigenvalues can be found in Appendix A. For eigenvectors, the first vector, with principal axes V_{12} and V_{22} , corresponds to column of the maximum eigenvalue Λ_{22} . The second vector, V_{11} and V_{21} , corresponds to column of minimum eigenvalue Λ_{11} .

The angle θ for the first eigenvector is measured against the x-axis, and is given by:

$$\theta_1 = \tan^{-1} \left(\frac{V_{22}}{V_{12}} \right) \quad (4-1)$$

And for the second eigenvector

$$\theta_2 = \tan^{-1} \left(\frac{V_{21}}{V_{11}} \right) \quad (4-2)$$

Clearly we have $\theta_2 = \theta_1 + 90^\circ$. The maximum eigenvalue together with the first eigenvector define the principal component (PC1).

Case 1: D1 and D2 are identical

Figure 4.13 shows the direction of PC1, and the associated second, orthogonal eigenvector, for the special case when $D1 = D2$. From Figure 4.13, it can be seen that all data points lie on PC1. In addition, the angle θ_1 is 45° , and one of the eigenvalues is zero and the eigenvalue Λ_{22} reaches maximum. All these are attributes for the two vectors being identical. Table 4.3 summarises the output parameters from performing PCA of $D1$ and $D2$.

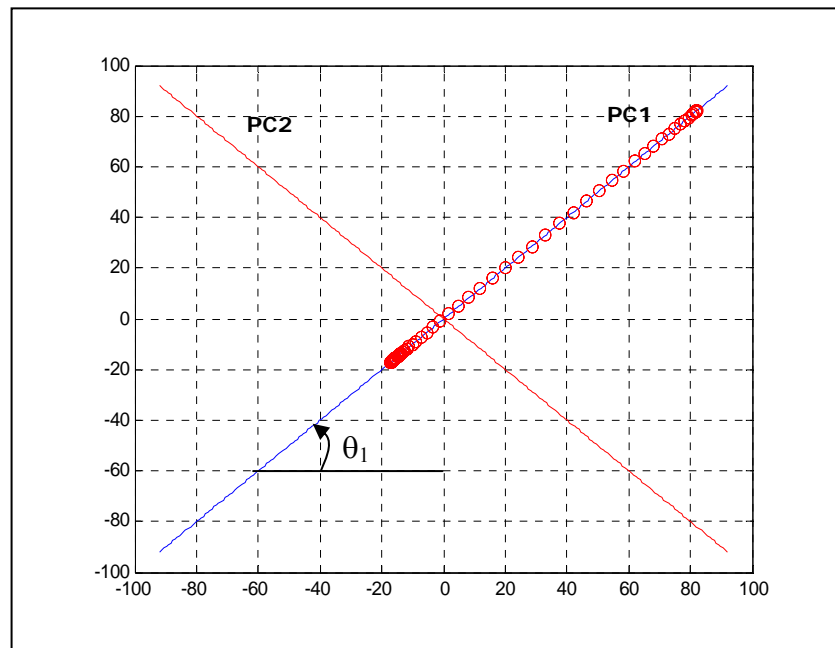


Figure 4.13 A principal component (PC1) and second component (PC2)

Table 4.3 Result of PCA

| V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.7071 | 0.7071 | 1977.1 | 0 | 45 | 1 | 0 |

Case 2: $D2$ shifted by Δ

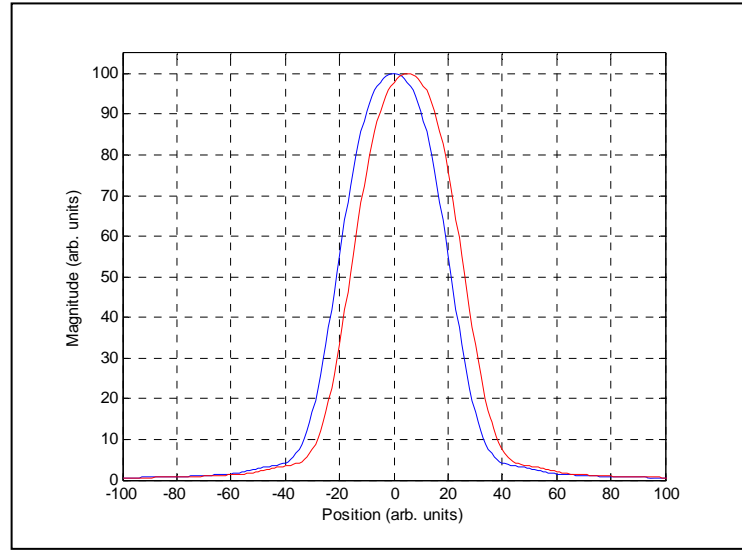


Figure 4.14 $D1$, blue, and $D2$, red, shifted by Δ

In this case $D1$ and $D2$ are identical but $D2$ is shifted laterally by an amount of Δ . Figure 4.15 shows the data plot and the eigenvectors. As can be seen, the angle θ_1 remains 45° although the data points are spreading away from the eigenvector. This point indicates that the minimum eigenvalue is not zero. Indeed changing the shift Δ will only change the eigenvalues but not the directions of the eigenvectors. Table 4.4 shows the resulting parameters as Δ is increased. Note that the last column in the table, $\Lambda_{11}/\Lambda_{22}$, is related to the shift Δ in a parabolic fashion.

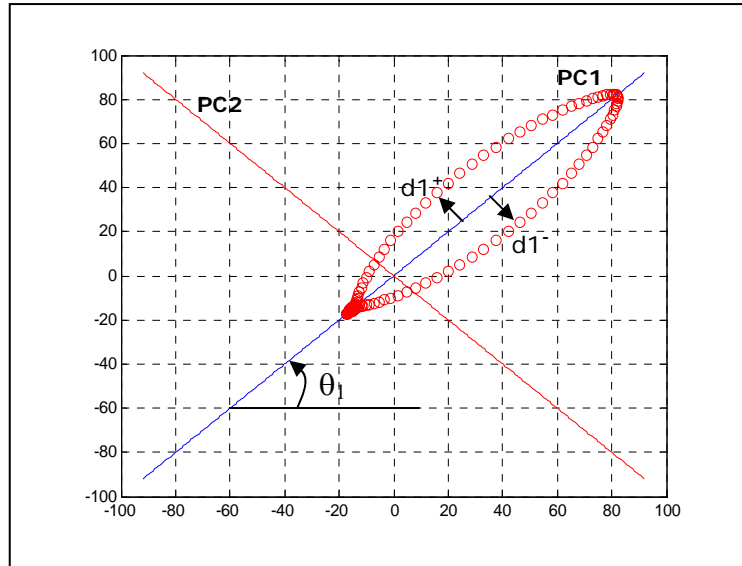


Figure 4.15 Data points on the PCA axis

Table 4.4 Result of PCA

| Δ (pixels) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0 | 0.7071 | 0.7071 | 1977.1 | 0 | 45 | 1 | 0 |
| 1 | 0.7071 | 0.7071 | 1975.9 | 1.23 | 45 | 1 | 6.26E-4 |
| 2 | 0.7071 | 0.7071 | 1972.1 | 4.94 | 45 | 1 | 2.50E-3 |
| 3 | 0.7071 | 0.7071 | 1966.0 | 11.09 | 45 | 1 | 5.64E-3 |
| 4 | 0.7071 | 0.7071 | 1957.4 | 19.64 | 45 | 1 | 1.00E-2 |
| 5 | 0.7071 | 0.7071 | 1946.6 | 30.53 | 45 | 1 | 1.56E-2 |
| 10 | 0.7071 | 0.7071 | 1859.9 | 117.14 | 45 | 1 | 6.29E-2 |

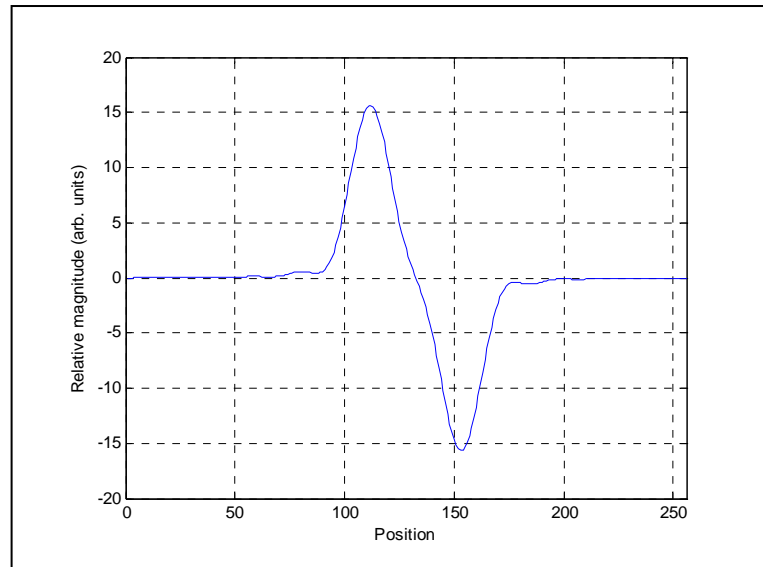


Figure 4.16 Graph of distance **d1**

Referring back to figure 4.15, we use **d1** to represent the distances of the data points from the eigenvector, and in figure 4.16, we plot **d1** as a function of the locations of the data points, showing the positive half (**d1**⁺) first followed by the negative half (**d1**⁻). As can be seen the positive and the negative parts are symmetrical about the eigenvector. If we calculate the standard deviation of **d1** over the positive half (or negative), and tabulate the results against the shift Δ , Table 4.5 results. As can be seen the values of the STD(**d1**) relate closely to those of Δ .

Table 4.5 Standard deviation (STD) of $\mathbf{d1}^+$ ($\mathbf{d1}^-$)

| Shifted Δ (pixels) | STD of $\mathbf{d1}^+$ ($\mathbf{d1}^-$) | Difference |
|---------------------------|--|------------|
| 0 | 0 | 0 |
| 1 | 0.96 | 0.04 |
| 2 | 1.92 | 0.08 |
| 3 | 2.87 | 0.13 |
| 4 | 3.83 | 0.17 |
| 5 | 4.76 | 0.24 |
| 6 | 5.70 | 0.30 |
| 7 | 6.61 | 0.39 |
| 8 | 7.53 | 0.47 |
| 9 | 8.40 | 0.60 |
| 10 | 9.30 | 0.70 |

The $\text{STD}(\mathbf{d1})$ can serve as a check of the closeness between $D1$ and $D2$. If we measure a $\text{STD}(\mathbf{d1}) = \mathbf{a}$, we can use it to shift the data set $D2$ by a distance \mathbf{a} , and then performing the PCA on $D1$ and the shifted version of $D2$ a second time. A more accurate comparison between the two can be obtained. The shifting of $D2$ is accomplished by using the shift property of the Fourier Transform, which is shown Equation 4-3 and 4-4.

$$G(f) = \mathfrak{F}\{g(t)\} \quad (4-3)$$

$$\mathfrak{F}\{g(t-a)\} = e^{-i2\pi fa} G(f) \quad (4-4)$$

The examples of performing PCA before and after shifting $D2$ are shown in Figure 4.17 and 4.18 respectively.

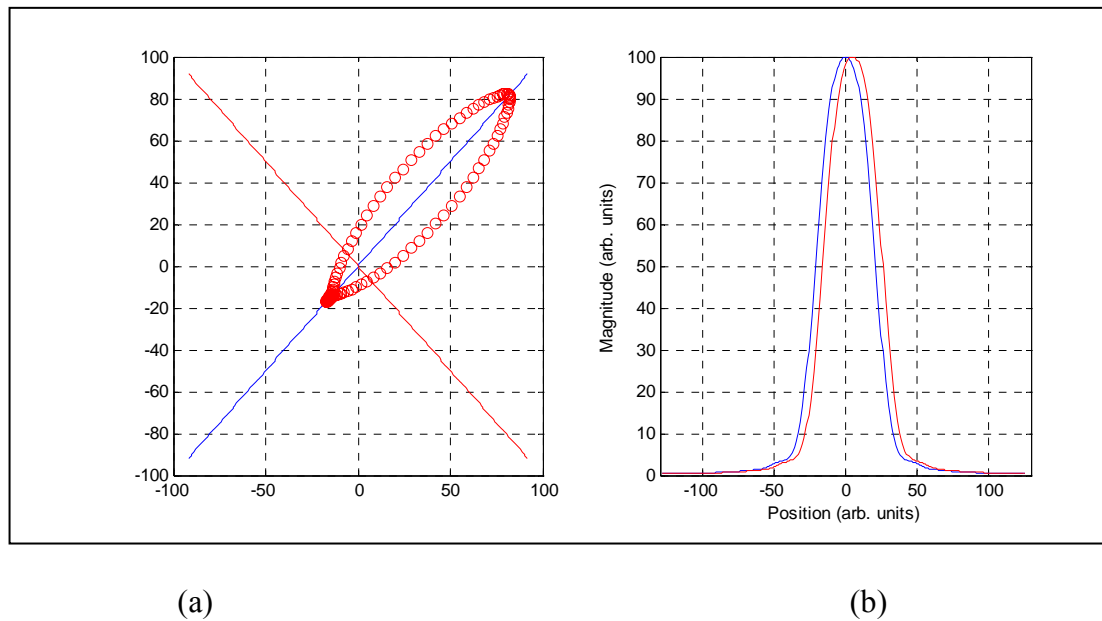


Figure 4.17 Data before shift $D2$; (a) data points on the PCA axis, (b) $D1$ (blue) $D2$ (red)

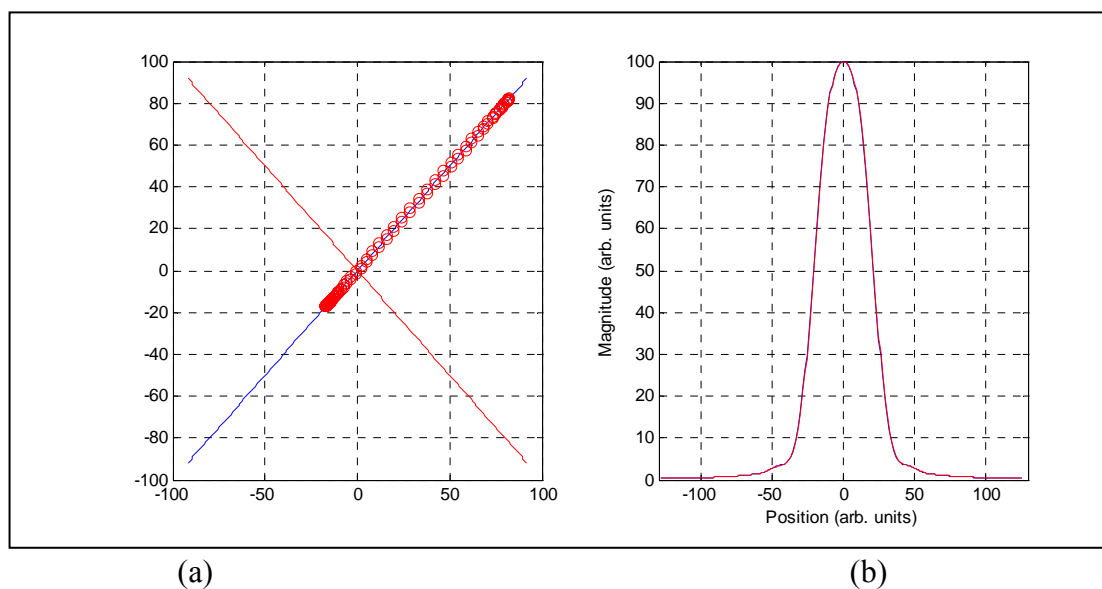


Figure 4.18 Data after shift $D2$ by $\Delta = 4.76$; (a) data points on the PCA axis, (b) $D1$ (blue) and shifted $D2$ (red)

Case 3: The width of $D1$ and $D2$ are different

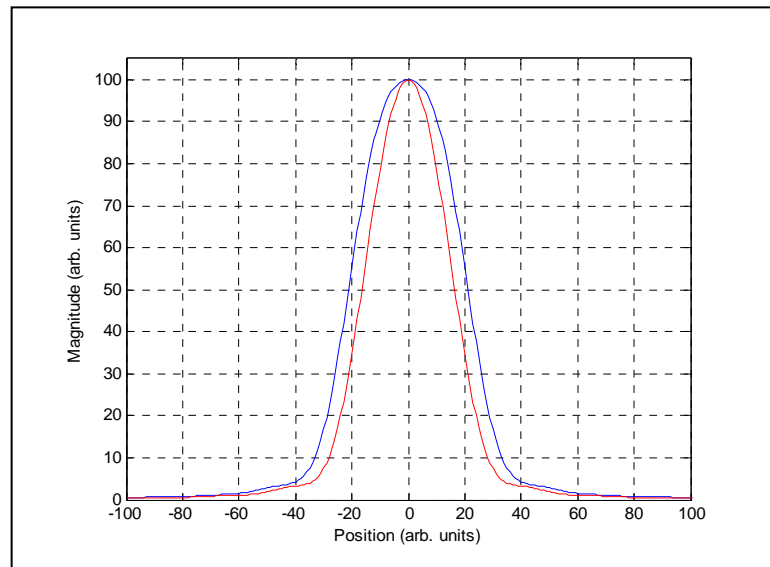


Figure 4.19 Graph $D1$, blue, and $D2$, red

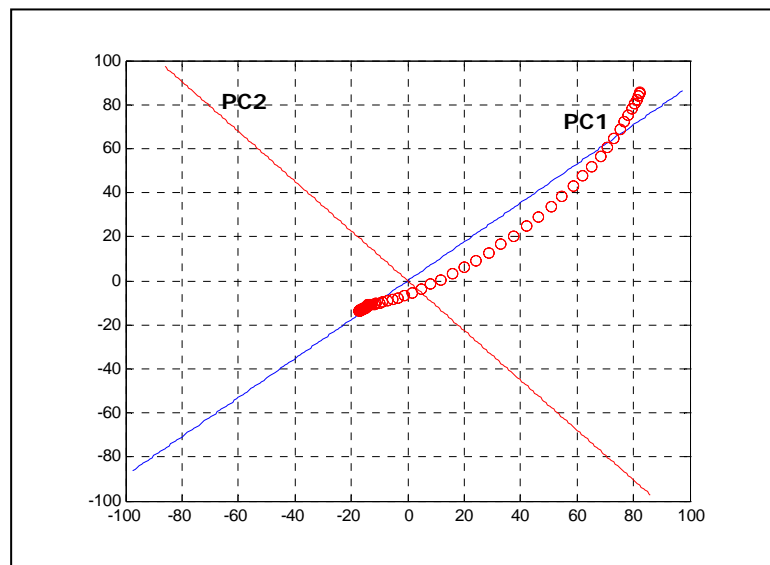


Figure 4.20 The distribution of data points on the PCA axis

In this case the width of $D2$ is smaller than $D1$ (see figure 4.19), the width of $D1$ and $D2$ were measured by full width at half max (FWHM) method. Figure 4.20 shows

the distribution of data points from performing PCA. The data points are single value function and do not lie on PC1.

Table 4.6 Result of PCA

| Width Ratio D1/D2 | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.619 | 0.6206 | 0.7841 | 1586.2 | 42.208 | 51.6 | 0.79 | 2.66E-2 |
| 0.714 | 0.6537 | 0.7567 | 1717.4 | 18.403 | 49.2 | 0.86 | 1.07E-2 |
| 0.762 | 0.6626 | 0.7489 | 1757.2 | 13.285 | 48.5 | 0.88 | 7.56E-3 |
| 0.857 | 0.6807 | 0.7324 | 1843.4 | 5.110 | 47.1 | 0.93 | 2.77E-3 |
| 0.905 | 0.6897 | 0.7240 | 1889.1 | 2.334 | 46.4 | 0.95 | 1.23E-3 |
| 0.952 | 0.6985 | 0.7155 | 1935.8 | 0.597 | 45.7 | 0.97 | 3.08E-4 |
| 1:1 | 0.7071 | 0.7071 | 1982.9 | 0 | 45 | 1 | 0 |
| 1.050 | 0.7155 | 0.6985 | 1935.8 | 0.597 | 44.3 | 1.02 | 3.08E-4 |
| 1.105 | 0.7240 | 0.6897 | 1889.1 | 2.334 | 43.6 | 1.05 | 1.23E-3 |
| 1.235 | 0.7324 | 0.6807 | 1843.4 | 5.110 | 42.9 | 1.07 | 2.77E-3 |
| 1.312 | 0.7489 | 0.6626 | 1757.2 | 13.285 | 41.5 | 1.13 | 7.56E-3 |
| 1.400 | 0.7567 | 0.6537 | 1717.4 | 18.403 | 40.8 | 1.16 | 1.07E-2 |
| 1.616 | 0.7841 | 0.6206 | 1586.2 | 42.208 | 38.4 | 1.26 | 2.66E-2 |

From Table 4.6 it can be seen that all parameters of PCA are changed when the width of $D2$ changes. However the angle θ_1 tell us that the width of $D1$ is wider or

narrower than $D2$. If θ_1 is greater than 45° , the width of $D1$ is smaller than $D2$. In contrast if θ_1 is less than 45° , the width of $D1$ is bigger than $D2$.

Table 4.7 shows the tabulated values of the width ratio ($D1/D2$) against the eigenvectors components ratio (V_{12}/V_{22}). Based on the measured eigenvector ratio and table 4.7, a different template ($D1$) can be selected to perform PCA a second time, in order to produce a better matching of the two profiles.

Table 4.7 Ratio of $D1/D2$ and V_{12}/V_{22}

| $D1/D2$ | V_{12}/V_{22} | Difference |
|---------|-----------------|------------|
| 0.619 | 0.79 | 0.171 |
| 0.714 | 0.86 | 0.146 |
| 0.762 | 0.88 | 0.118 |
| 0.810 | 0.93 | 0.120 |
| 0.905 | 0.95 | 0.045 |
| 0.952 | 0.97 | 0.018 |
| 1 | 1 | 0 |
| 1.050 | 1.02 | 0.030 |
| 1.105 | 1.05 | 0.055 |
| 1.235 | 1.07 | 0.097 |
| 1.312 | 1.13 | 0.183 |
| 1.400 | 1.16 | 0.240 |
| 1.616 | 1.26 | 0.356 |

Case 4: $D2$ multiplied by factor a

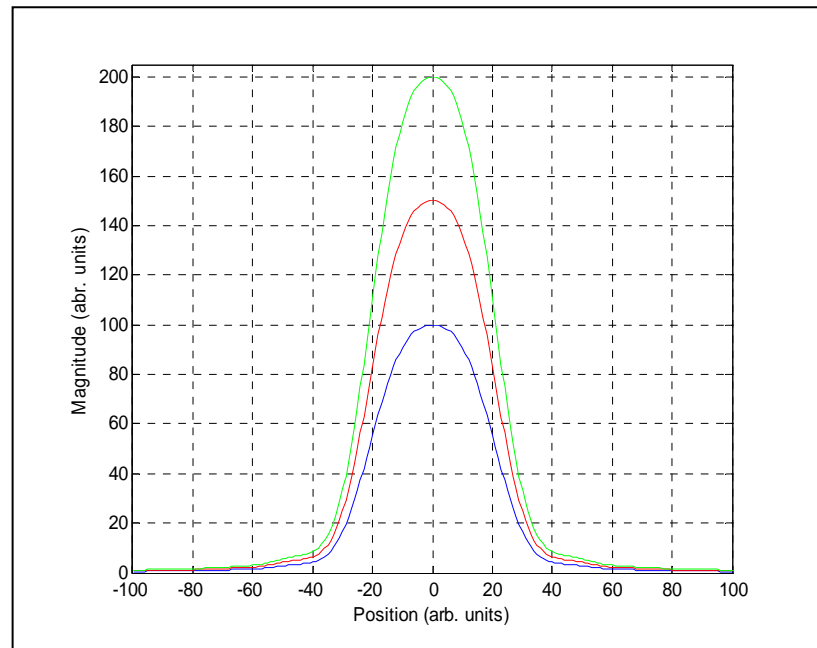


Figure 4.21 Graph of $D1$, blue, and $a \cdot D2$, red and green

In this case the data $D2$ is multiplied by factor a (see figure 4.21). Figure 4.22 shows the distribution of data points from performing PCA. It can be seen that data points lie on PC1.

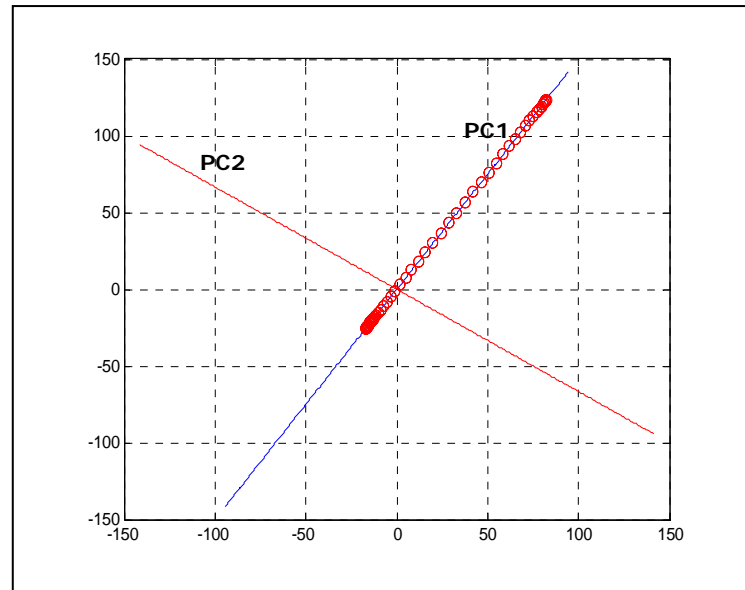


Figure 4.22 Data points on the PCA axis

Table 4.8 Result of PCA

| Factor a | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{22}/V_{12} | $\Lambda_{11}/\Lambda_{22}$ |
|------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.2 | 0.9805 | 0.1961 | 1031.1 | 0 | 11.31 | 0.2 | 0 |
| 0.3333 | 0.9487 | 0.3162 | 1101.6 | 0 | 18.43 | 0.333 | 0 |
| 0.4 | 0.9285 | 0.3714 | 1150.1 | 0 | 21.80 | 0.4 | 0 |
| 0.5 | 0.8944 | 0.4472 | 1239.3 | 0 | 26.56 | 0.5 | 0 |
| 0.6666 | 0.8320 | 0.5547 | 1432.1 | 0 | 33.69 | 0.666 | 0 |
| 1 | 0.7071 | 0.7071 | 1982.9 | 0 | 45 | 1 | 0 |
| 1.5 | 0.5547 | 0.8320 | 3222.1 | 0 | 56.31 | 1.5 | 0 |
| 2 | 0.4472 | 0.8944 | 4957.1 | 0 | 63.43 | 2 | 0 |
| 2.5 | 0.3714 | 0.9285 | 7187.9 | 0 | 68.20 | 2.5 | 0 |
| 3 | 0.3162 | 0.9487 | 9914.3 | 0 | 71.56 | 3 | 0 |
| 5 | 0.1961 | 0.9805 | 25777 | 0 | 78.70 | 5 | 0 |

The Table 4.8 shows that maximum eigenvalue Λ_{22} and angle θ_1 vary according to \mathbf{a} . However, the minimum eigenvalue Λ_{11} of all factors of \mathbf{a} are still zero. In conclusion, if eigenvalues Λ_{11} of PCA is zero and angle θ_1 is not 45° , it means data set $D2$ is multiplied by factor \mathbf{a} .

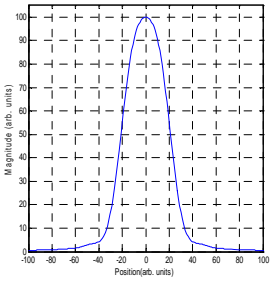
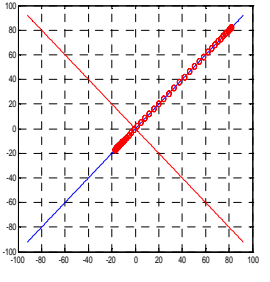
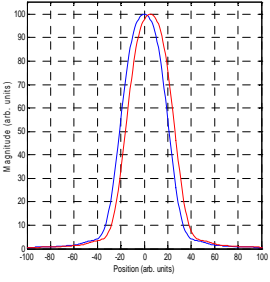
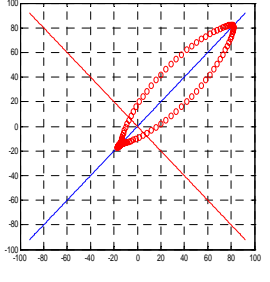
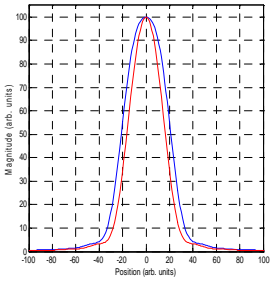
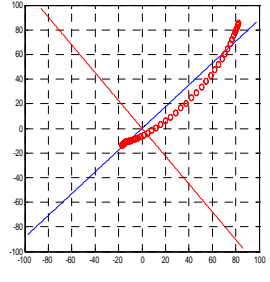
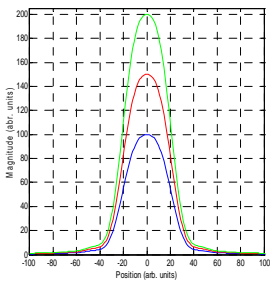
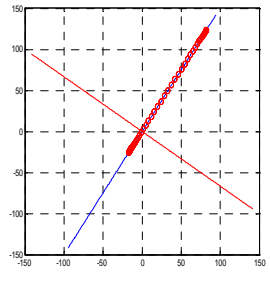
The factor \mathbf{a} can be simply calculated from the ratio of eigenvectors.

$$a = \frac{V_{22}}{V_{12}} \quad (4-5)$$

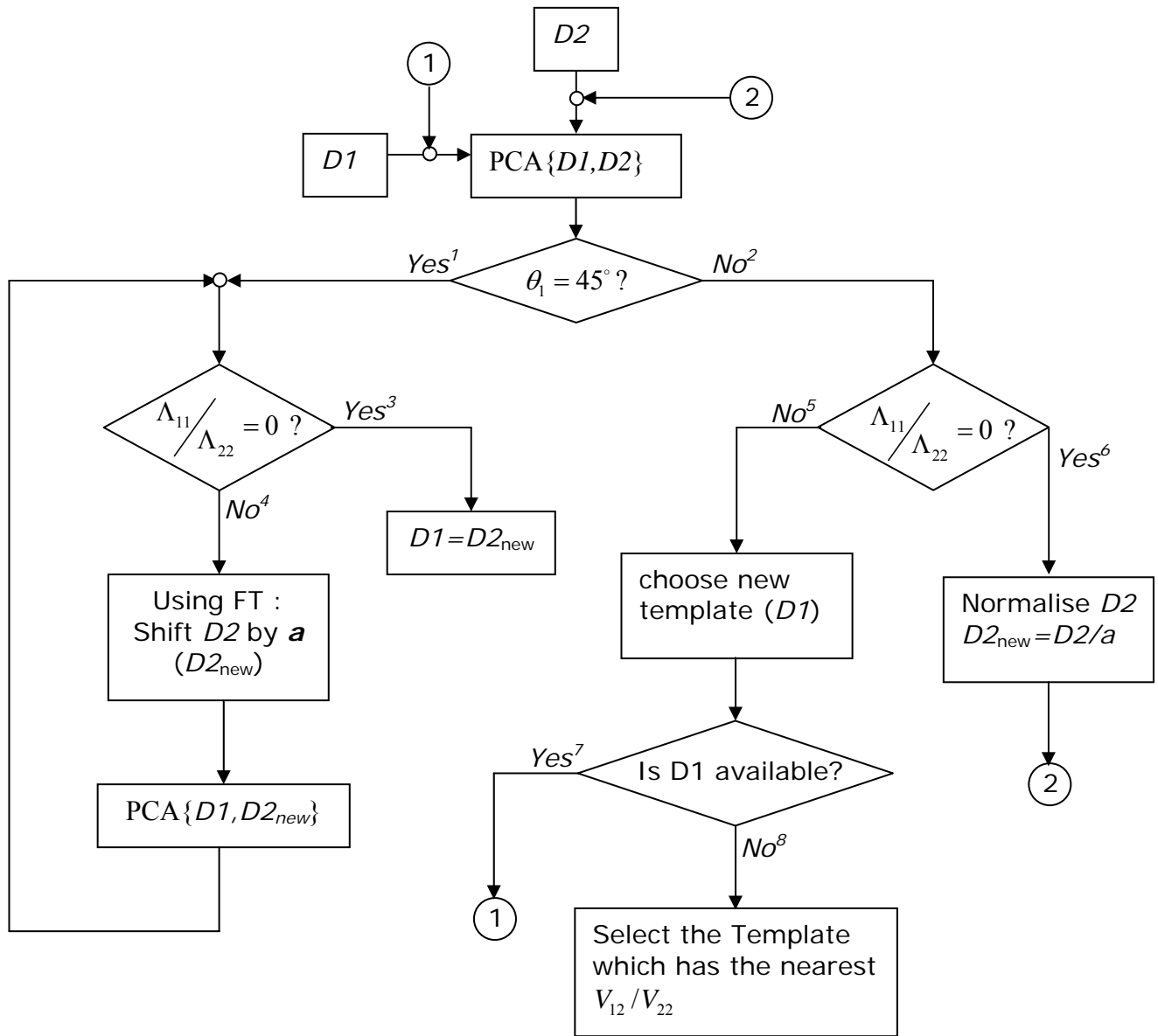
It should be note that this section is for completion only, in reality the signals are normalised.

The distribution of data points plot on PCA axis for each case and its characteristics are shown in Table 4.9.

Table 4.9 Performing PCA for each case of data set using profiles of images

| Case/Characteristics | Graph of data (image profiles) | Distribution of data points on PCA axis |
|--|---|---|
| 1. Data set $D1=D2$ $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} = 0$ |  |  |
| 2. Data set $D2$ shifted by Δ $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} \neq 0$ |  |  |
| 3. The width of $D1$ (blue) and $D2$ (red) are different $\theta_1 \neq 45^\circ$ $\Lambda_{11}/\Lambda_{22} \neq 0$ |  |  |
| 4. $D2$ multiplied by factor a $\theta_1 \neq 45^\circ$ $\Lambda_{11}/\Lambda_{22} = 0$ |  |  |

The flow chart below shows the whole measurement procedure.



The meaning of each superscript in the flow chart is described as following.

¹Data set $D1$ and $D2$ are the same size

²Data set $D1$ and $D2$ are different (magnitude/width)

³Data set $D2$ is not shifted ($D1=D2$)

⁴Data set $D2$ is shifted

⁵Data set $D1$ and $D2$ are not the same width

⁶Data set $D1$ and $D2$ are the same width but different in magnitude

⁷Template $D1$ is available

⁸Template $D1$ is not available

For each time when performing PCA, if the two data sets are not perfect match, $\theta_1 \neq 45^\circ$ and $\Lambda_{11} / \Lambda_{22} \neq 0$, we can correct the data set by using the suitable technique for each case as described above. After that repeat performing PCA until the system obtains the perfect match.

4.6.2 Add noise to system

The following example illustrates the effect of noise to system when the different noise levels are added. We will first use a small amount of data points to examine the change of the PCA parameters. The data set $D1$ has 15 data points. Data set $D2$ is formed by adding different percentages of normally distributed random noise to $D1$. Both data sets are in Table 4.10.

Table 4.10 Data set *D1* and *D2*

| <i>D1</i> | <i>D2</i> with noise (%) | | | | | | | | | |
|-----------|--------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 0 | 1 | 2 | 3 | 5 | 7 | 10 | 13 | 16 | 20 |
| 0.9 | 0.9 | 0.90 | 0.91 | 0.94 | 0.90 | 0.74 | 0.89 | 0.97 | 0.75 | 0.69 |
| 1 | 1 | 1.00 | 1.02 | 0.94 | 1.01 | 1.08 | 1.20 | 1.08 | 1.07 | 0.61 |
| -1 | -1 | -1.01 | -1.03 | -0.99 | -1.03 | -1.13 | -1.19 | -0.92 | -0.99 | -0.89 |
| 2 | 2 | 2.01 | 1.96 | 1.88 | 2.07 | 1.96 | 1.93 | 1.88 | 2.13 | 1.52 |
| 0.5 | 0.5 | 0.50 | 0.50 | 0.49 | 0.50 | 0.54 | 0.52 | 0.49 | 0.61 | 0.54 |
| 1.6 | 1.6 | 1.59 | 1.63 | 1.60 | 1.66 | 1.60 | 1.77 | 1.93 | 1.08 | 1.67 |
| -11 | -11 | -11.1 | -11.6 | -11.1 | -10.9 | -11.5 | -11.1 | -13.0 | -9.7 | -10.4 |
| 13 | 13 | 12.87 | 12.51 | 12.74 | 12.57 | 12.07 | 11.58 | 14.14 | 15.45 | 10.45 |
| -6 | -6 | -6.03 | -6.11 | -5.82 | -5.64 | -6.26 | -6.15 | -4.67 | -5.78 | -6.84 |
| -1.7 | -1.7 | -1.68 | -1.68 | -1.80 | -1.68 | -1.78 | -1.67 | -1.49 | -1.40 | -1.85 |
| 13 | 13 | 13.13 | 12.79 | 13.17 | 12.24 | 12.10 | 13.36 | 11.12 | 12.64 | 14.01 |
| 8 | 8 | 7.91 | 7.66 | 8.21 | 8.04 | 8.17 | 7.23 | 7.34 | 7.60 | 6.39 |
| 28 | 28 | 27.69 | 28.07 | 29.19 | 27.79 | 27.76 | 32.11 | 27.69 | 29.50 | 27.23 |
| 15 | 15 | 14.93 | 14.77 | 15.76 | 15.59 | 13.91 | 15.26 | 18.75 | 10.17 | 12.41 |
| 4 | 4 | 4.01 | 3.96 | 3.83 | 4.09 | 4.29 | 3.70 | 3.75 | 3.88 | 3.97 |

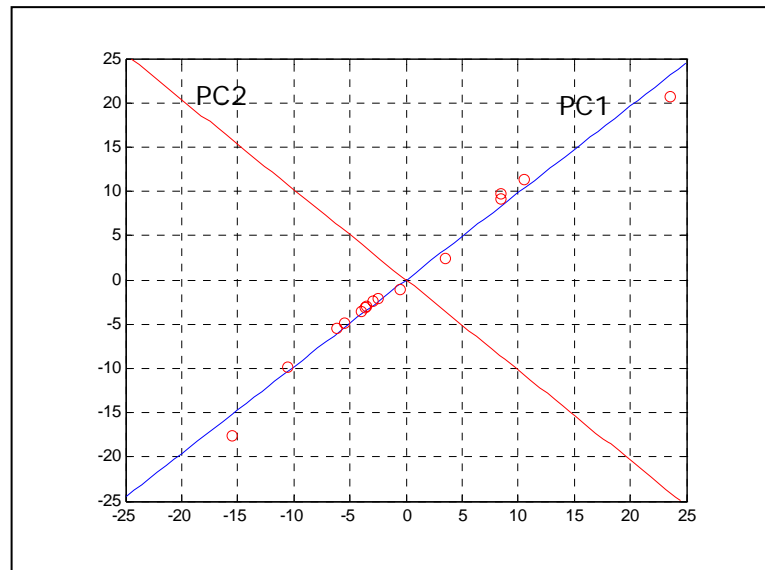


Figure 4.23 The distribution of data points on the PCA axis with noise =10%

From Figure 4.23, the distribution of data points is spread about PC1. They will be spread further when the noise level is increased.

Table 4.11 Result of PCA each row is the average of 100 simulations, each with a different noise pattern

| noise (%) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-----------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0 | 0.7071 | 0.7071 | 184.94 | 0 | 45.00 | 1.0000 | 0 |
| 1 | 0.7071 | 0.7071 | 184.95 | 4.2E-03 | 45.00 | 1.0000 | 2.3E-05 |
| 2 | 0.7072 | 0.7070 | 184.89 | 1.7E-02 | 44.99 | 1.0004 | 9.1E-05 |
| 3 | 0.7070 | 0.7071 | 184.98 | 3.7E-02 | 45.00 | 1.0000 | 2.0E-04 |
| 5 | 0.7070 | 0.7071 | 185.03 | 9.6E-02 | 45.00 | 1.0000 | 5.2E-04 |
| 7 | 0.7069 | 0.7071 | 185.03 | 1.9E-01 | 45.01 | 1.0003 | 1.0E-03 |
| 10 | 0.7071 | 0.7067 | 184.85 | 3.8E-01 | 44.99 | 1.0016 | 2.1E-03 |
| 13 | 0.7067 | 0.7065 | 185.27 | 6.5E-01 | 44.99 | 1.0031 | 3.6E-03 |
| 16 | 0.7059 | 0.7070 | 185.73 | 8.6E-01 | 45.04 | 1.0021 | 4.7E-03 |
| 20 | 0.7057 | 0.7060 | 186.07 | 1.50 | 44.98 | 0.9995 | 8.7E-03 |

From Table 4.11 it can be seen that the angle θ_1 is about 45° . The ratio of eigenvalues, the last column of Table, increases exponentially with respect to the noise levels.

Now we will use profiles of the images as data sets. This means the number of data points is now increased to 256. A profile without noise (Figure 4.24) serves as data set $D1$. Then data set $D2$ is formed by adding normally distributed random shot noise to $D1$. Figure 4.25 shows graph of data set $D2$ with the SNR=100.

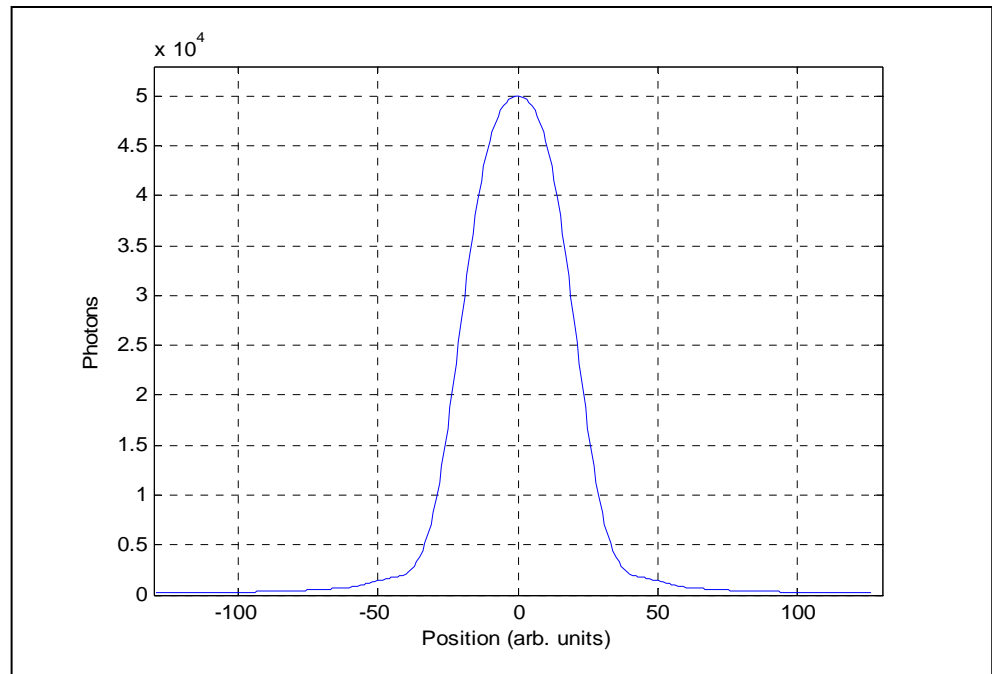


Figure 4.24 Profile of image without noises; $D1$

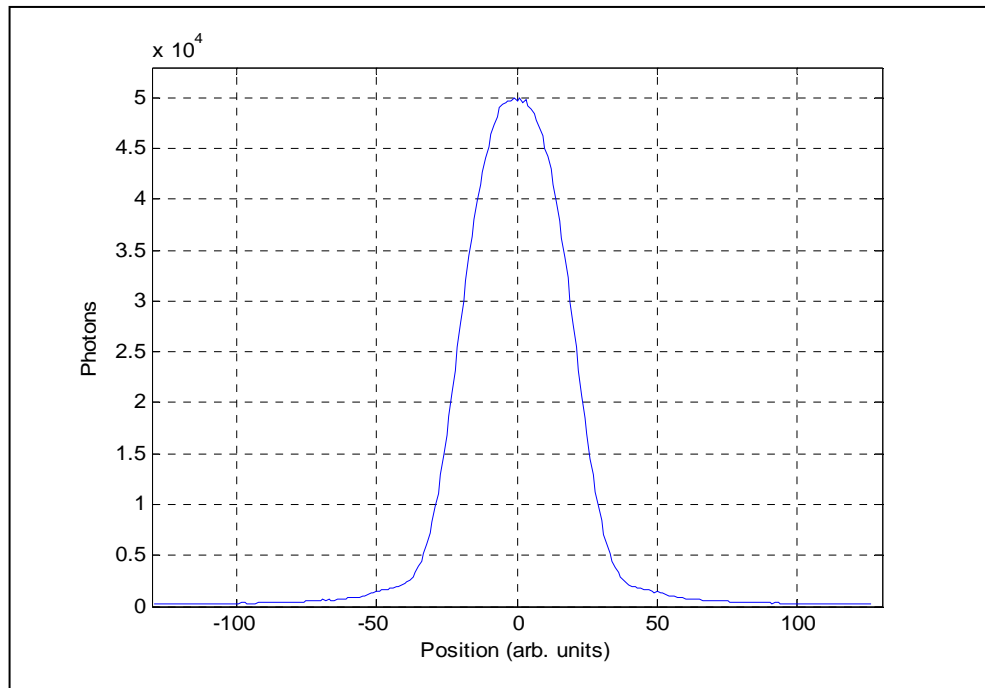


Figure 4.25 Profile of an image with noise; $D2$

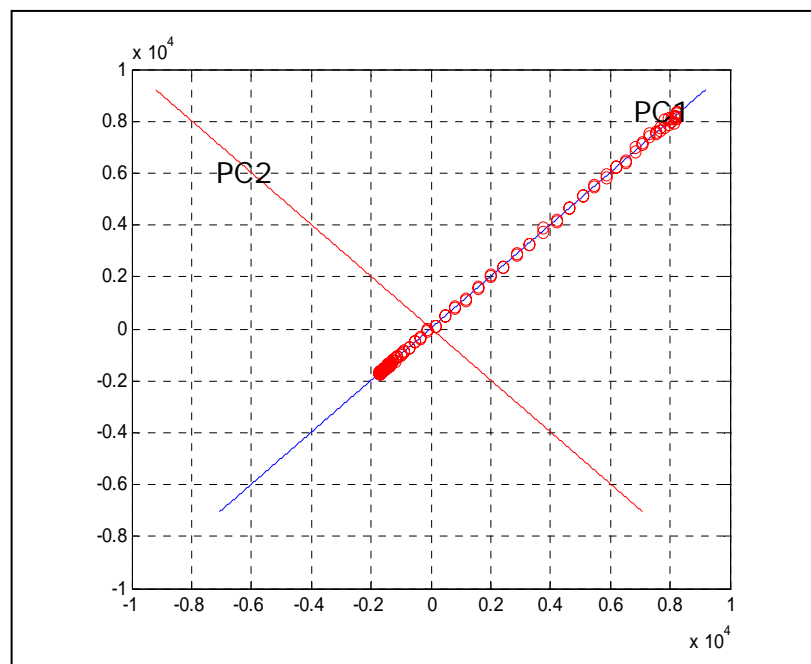


Figure 4.26 Data points on the PCA axis with SNR = 100

Table 4.12 Result of PCA

| SNR | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-----|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 10 | 0.7048 | 0.7093 | 1.89E3 | 8.63 | 45.18 | 0.993 | 4.34E-3 |
| 20 | 0.7063 | 0.7078 | 3.2E4 | 35.10 | 45.06 | 0.998 | 1.10E-3 |
| 30 | 0.7069 | 0.7072 | 1.6E5 | 77.98 | 45.013 | 0.999 | 4.85E-4 |
| 50 | 0.7069 | 0.7073 | 1.24E6 | 210.27 | 45.016 | 0.999 | 1.69E-4 |
| 80 | 0.7072 | 0.7069 | 8.12E6 | 541.39 | 44.988 | 0.999 | 6.66E-5 |
| 100 | 0.7071 | 0.7071 | 1.98E7 | 852.67 | 45 | 1 | 4.30E-5 |
| 150 | 0.7071 | 0.7071 | 1.00E8 | 1950.9 | 45 | 1 | 1.94E-5 |
| 200 | 0.7071 | 0.7071 | 3.17E8 | 3470.1 | 45 | 1 | 1.09E-5 |
| 223 | 0.7071 | 0.7071 | 4.95E8 | 4155.6 | 45 | 1 | 8.38E-6 |
| 316 | 0.7071 | 0.7071 | 1.98E9 | 8859.6 | 45 | 1 | 4.46E-6 |

From Table 4.12 it can be seen that adding noises to the system changes the eigenvalues, with the ratio $\Lambda_{11}/\Lambda_{22}$ decreases as SNR increases. Noise will also affect the angle θ_1 of the PC1, but only by very small amount.

Referring back to Table 4.4 where the effects of shifting $D2$ relative to $D1$ were tabulated in the absence of noise, shifting $D2$ by one pixel would change the eigenvalues ratio $\Lambda_{11}/\Lambda_{22}$ by approximately 6.26E-4. Comparing this value to those in Table 4.12, we can conclude that a SNR of around 60 would result in an uncertainty in the shift of the two profiles of about 1 pixel.

We further demonstrate in the case where the data set $D2$ would contain noises and is also shifted by Δ . An example of $D1$ and $D2$ shows in Figure 4.27. The results of performing PCA, with $\Delta=10$ pixels and different SNR, are shown in Table 4.13. From Table 4.13 we can see again the effect of noise on the angle θ_1 is very small. However, the change in the eigenvalues ratio is non-linearly related to the amount of shift, even for the sample values of SNR. As is obvious, when both $D1$ and $D2$ are noisy, the net effect is given by the summation, incoherently, of the two noise profiles.

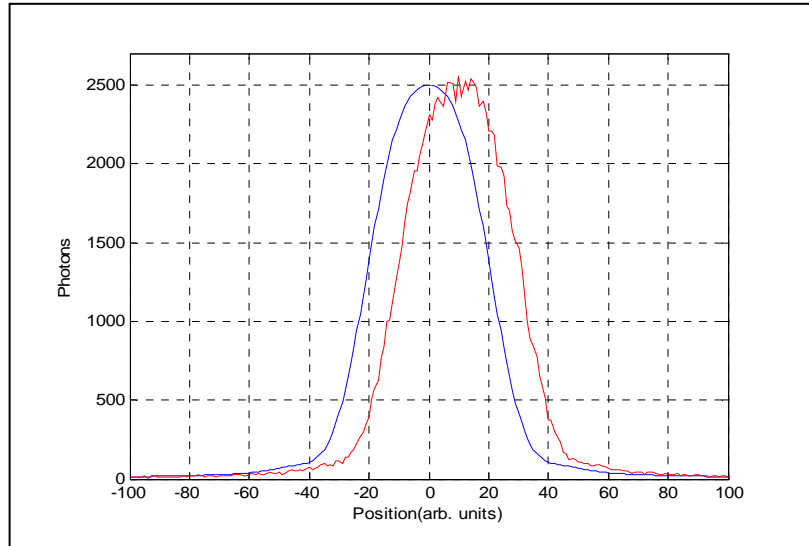


Figure 4.27 Data $D1$ (blue) and $D2$ (red) with $\Delta=10$ pixels and SNR=50.

Table 4.13 Result of PCA

| SNR | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-----|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 10 | 0.7062 | 0.7079 | 2.01E+03 | 1.86E+01 | 45.066 | 0.998 | 9.29E-03 |
| 20 | 0.7069 | 0.7073 | 3.17E+04 | 8.90E+01 | 45.017 | 0.999 | 2.80E-03 |
| 30 | 0.7070 | 0.7072 | 1.61E+05 | 2.58E+02 | 45.012 | 1.000 | 1.60E-03 |
| 50 | 0.7071 | 0.7071 | 1.24E+06 | 1.21E+03 | 45.002 | 1.000 | 9.75E-04 |
| 80 | 0.7072 | 0.7070 | 8.12E+06 | 6.10E+03 | 44.988 | 1.000 | 7.51E-04 |
| 100 | 0.7071 | 0.7071 | 1.98E+07 | 1.40E+04 | 45.003 | 1.000 | 7.08E-04 |
| 150 | 0.7071 | 0.7071 | 1.00E+08 | 6.60E+04 | 44.999 | 1.000 | 6.58E-04 |
| 200 | 0.7072 | 0.7070 | 3.17E+08 | 2.03E+05 | 44.994 | 1.000 | 6.40E-04 |

Thus, in conclusion, adding noise to the system only changes the eigenvalues ratio $\Lambda_{11}/\Lambda_{22}$, it does not affect the angle θ_1 significantly. This means that when performing PCA, if the results obtained are $\theta_1 \neq 45^\circ$ and eigenvalues $\Lambda_{11}/\Lambda_{22} \neq 0$, this implies no matching between the two. In this case a different template (*D1*) should be used until we obtain the PCA result of $\theta_1 = 45^\circ$. This would give two profiles of equal width. Techniques such as described in case 2 can then be used to minimise the eigenvalues ratio $\Lambda_{11}/\Lambda_{22}$, which would make *D2* approach *D1*.

Chapter 5

5 Two-Dimension PCA

In this chapter we will extend the PCA described before to two-dimension line segment images. Thus features of line segment such as lengths and shapes, which could not be measured using the line profiles, can now be determined. The use of PCA to measure these parameters from the line segment images will be described first. After that we will discuss the use of Hough transform. The purposes of HT are to provide initial information, mainly relating to location, orientation, shape, and rough dimensions of object, before applying 2D PCA to extract more accurate measurements.

5.1 Convert Line Image to Vector

In order to perform PCA of an image, the image firstly is converted into a column vector. A column vector can be constructed by concatenating each row with the previous row in sequence. As in Figure 5.1, a $N \times M$ image is converted into a single column $NM \times 1$ vector, the row R1 correspond to C1, R2 correspond to C2 and so on. Thus an image of 256x256 pixels is converted into a column vector of length 65536 x 1. Figure 5.2 shows an image with the associated vector in Figure 5.3a (shown as row for convenience). Figure 5.3b is a magnified version of (a) in red box, showing each small segment is in fact a profile along the length of the image.

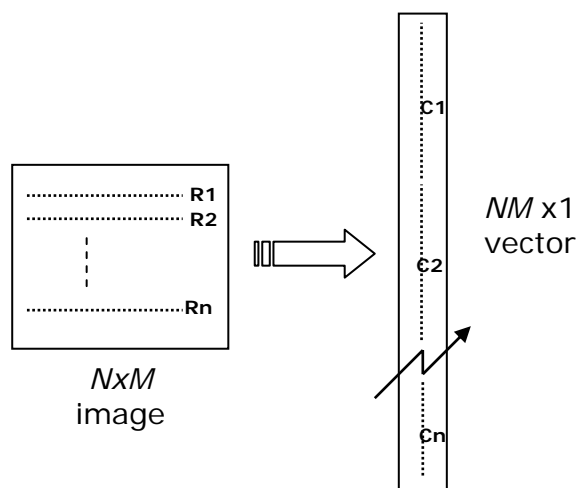


Figure 5.1 Convert image to column vector

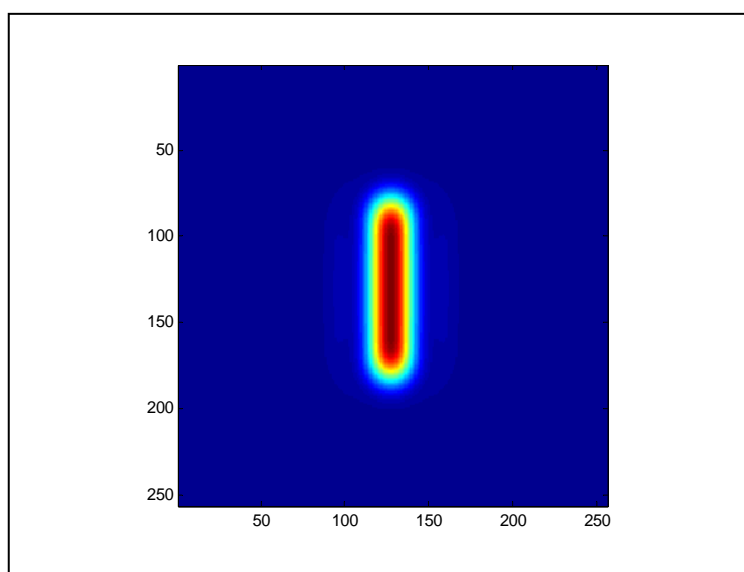


Figure 5.2 Figure of image

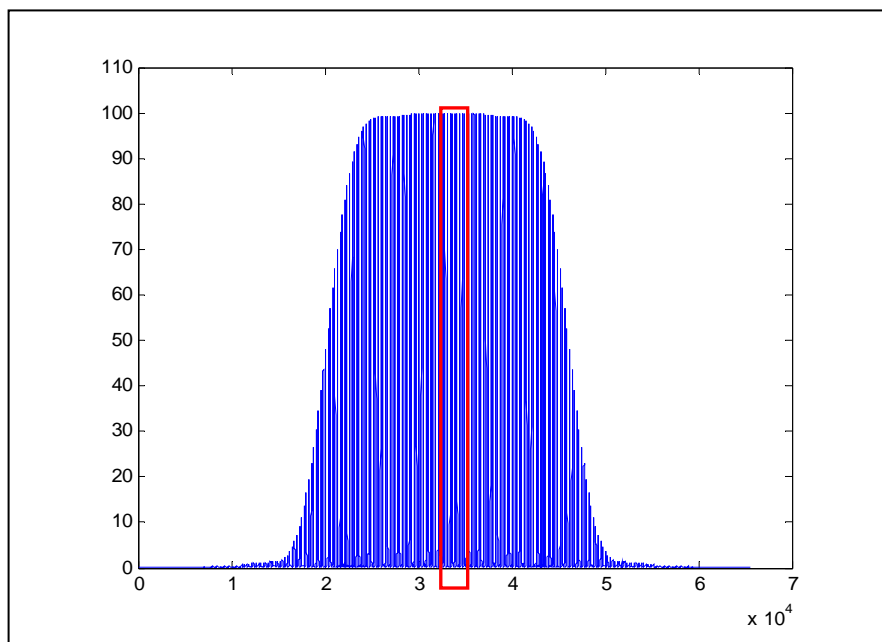


Figure 5.3a Graph of vector

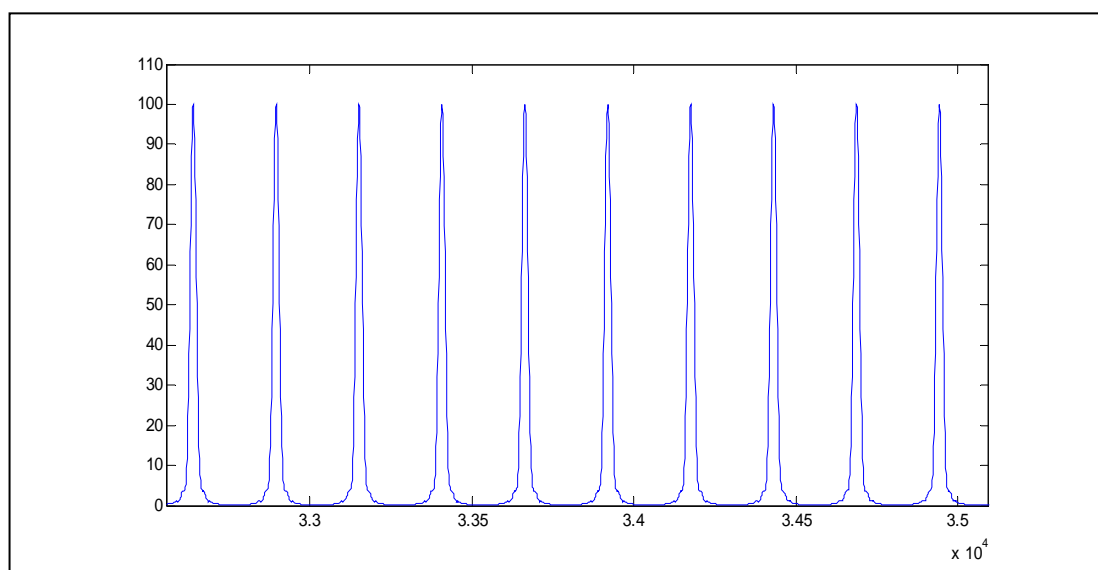


Figure 5.3b Magnified graph of vector in Figure 5.3a

5.2 Measuring line segment images

For simplicity we will consider the sample being a straight line segment, although the argument can be extended readily to other type of line images. I_2 is used to represent a known image which serves as the template; and I_1 , an unknown image obtained from the straight line segment whose characteristics we want to measure. Figure 5.4 and 5.5 show examples of I_1 and I_2 . We will examine the outcomes of performing PCA with different templates I_1 .

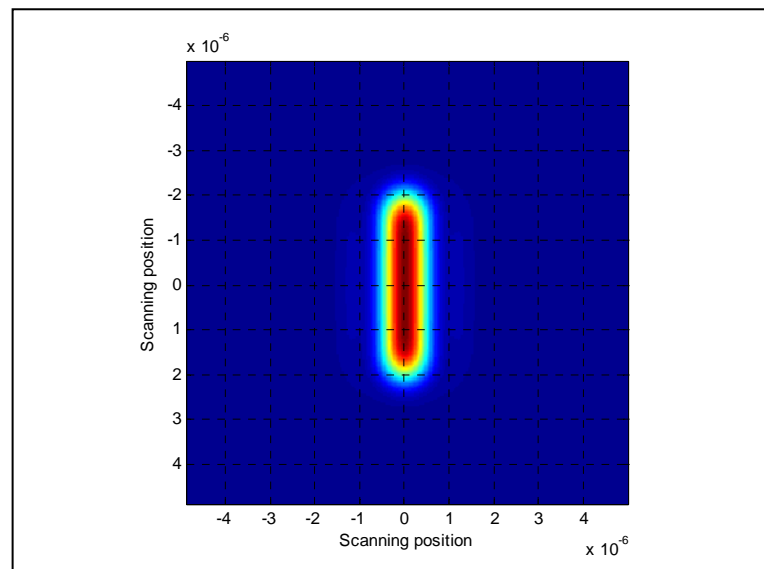


Figure 5.4 Line segment image I_1 (Line object: width= 500nm,
Length=4 μm)

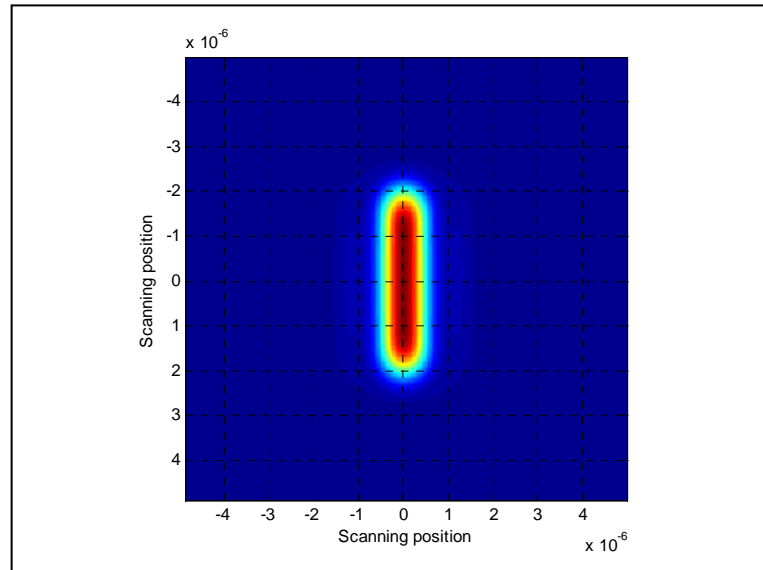


Figure 5.5 Line segment image I_2

Case 1: I_1 and I_2 are identical

Figure 5.6 shows the direction of the principal component PC1, and the associated second, orthogonal component, for the special case when $I_1 = I_2$. From Figure 5.6, it can be seen that all data points lie on PC1. In addition, the angle θ_1 is 45° , and one of the eigenvalues is zero and the eigenvalue Λ_{22} reaches maximum. All these are attributes for the two images being identical. Table 5.1 summarises the output parameters from performing PCA of I_1 and I_2 .

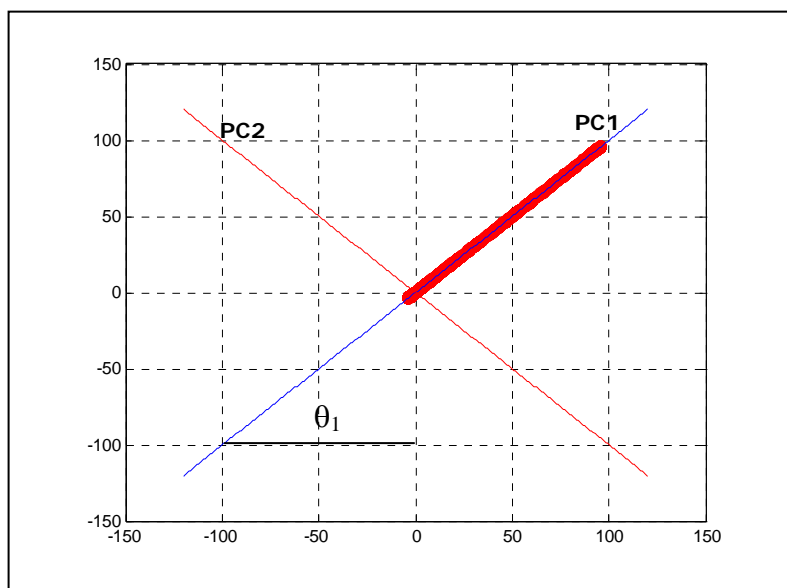


Figure 5.6 A principal component (PC1) and second component (PC2)

Table 5.1 Result of PCA

| V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.7071 | 0.7071 | 433.02 | 0 | 45.00 | 1.00 | 0 |

Case 2: I_2 shifted by Δx and/or Δy

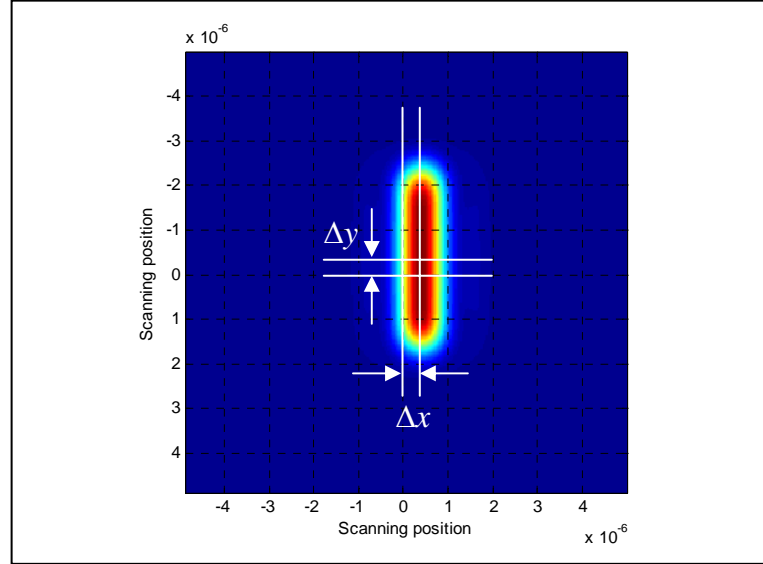


Figure 5.7 Image I_2 shifted by $\Delta x = 10$ pixels and $\Delta y = -10$ pixels

In this case I_1 and I_2 are identical but I_2 is shifted laterally by an amount of Δx and/or Δy , see Figure 5.7. Figure 5.8 shows the data plot and the eigenvectors. As can be seen, the angle θ_1 remains 45° although the data points are spreading away from the eigenvector. This point indicates that the minimum eigenvalue is not zero. Indeed changing the shift Δ will only change the eigenvalues, as shown in the last column of Table 5.2, but not the directions of the eigenvectors. Table 5.2 shows the resulting parameters as Δx is increased, $\Delta y = 1$ pixels. In Figure 5.9, we plot the eigenvalues ratio as a function of the shift Δx . Each graph represents a different shift Δy . It can be seen the change of each graph follows the same fashion but the values of eigenvalues increase according to the amount of shift Δy .

Table 5.2 Result of PCA ; $\Delta y = 1$ pixels

| Δx (pixels) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0 | 0.7071 | 0.7071 | 432.93 | 0.08 | 45.00 | 1.00 | 1.93E-04 |
| 1 | 0.7071 | 0.7071 | 432.26 | 0.75 | 45.00 | 1.00 | 1.74E-03 |
| 2 | 0.7071 | 0.7071 | 430.26 | 2.75 | 45.00 | 1.00 | 6.40E-03 |
| 3 | 0.7071 | 0.7071 | 426.97 | 6.05 | 45.00 | 1.00 | 1.42E-02 |
| 4 | 0.7071 | 0.7071 | 422.44 | 10.57 | 45.00 | 1.00 | 2.50E-02 |
| 5 | 0.7071 | 0.7071 | 416.76 | 16.26 | 45.00 | 1.00 | 3.90E-02 |
| 6 | 0.7071 | 0.7071 | 410.02 | 23.00 | 45.00 | 1.00 | 5.61E-02 |
| 7 | 0.7071 | 0.7071 | 402.34 | 30.68 | 45.00 | 1.00 | 7.62E-02 |
| 8 | 0.7071 | 0.7071 | 393.84 | 39.17 | 45.00 | 1.00 | 9.95E-02 |
| 9 | 0.7071 | 0.7071 | 384.67 | 48.34 | 45.00 | 1.00 | 1.26E-01 |
| 10 | 0.7071 | 0.7071 | 374.97 | 58.04 | 45.00 | 1.00 | 1.55E-01 |

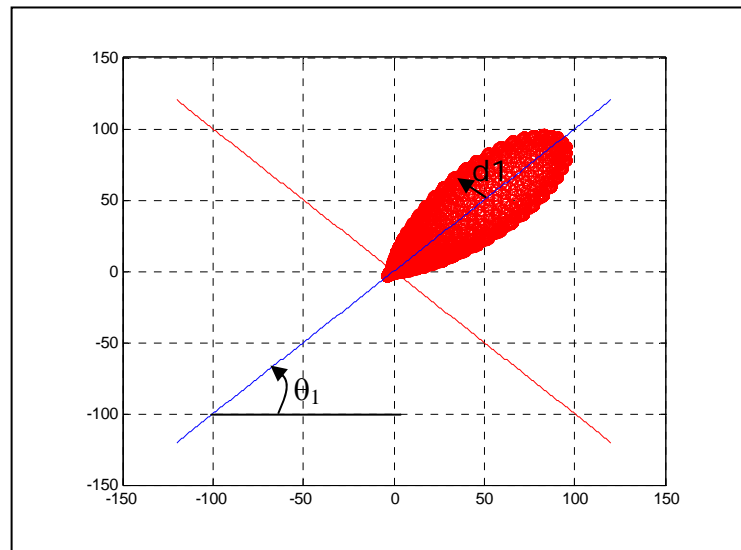


Figure 5.8 Data points on the PCA axis

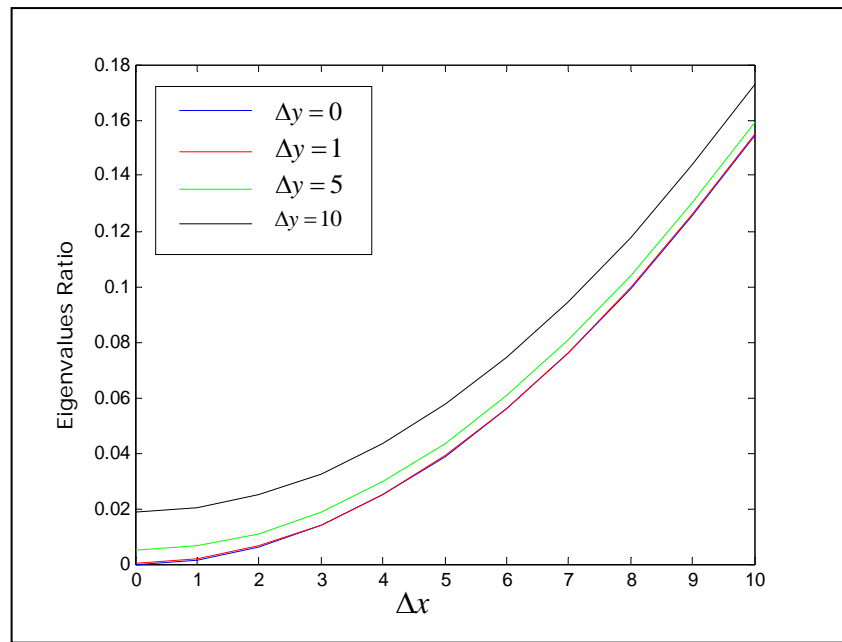


Figure 5.9 Eigenvalues ratio $\Lambda_{11} / \Lambda_{22}$

From Figure 5.8, we use **d1** to represent the distance of the data points from the eigenvector, and in Figure 5.10, we plot **d1** as a function of the locations of the data points, showing the positive and negative sides. As can be seen the positive and the negative parts are symmetrical about the eigenvector. Based on figures 5.8 and 5.10, we can shift the template in x-y directions until the value d1 is negligible.

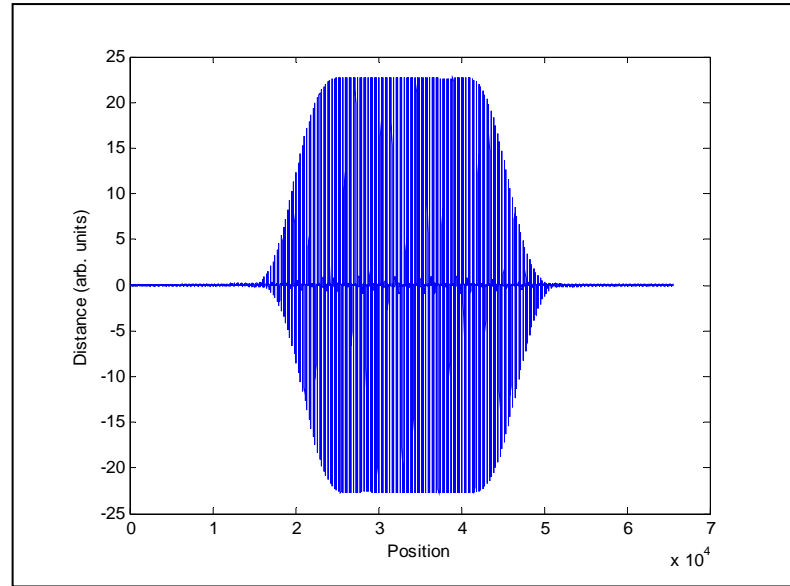


Figure 5.10 Graph of distance d_1 with $\Delta x = 5$ pixels and $\Delta y = 1$ pixel

Case 3: I_2 rotated by an angle θ

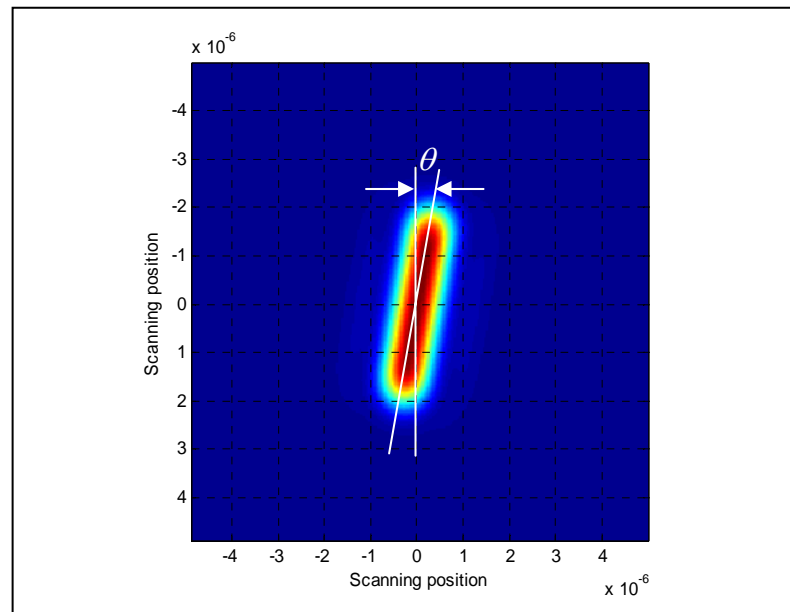


Figure 5.11 I_2 rotated by $\theta = 10^\circ$

Table 5.3 Result of PCA

| $I2$ Rotate θ° | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 1 | 0.7071 | 0.7070 | 432.71 | 0.18 | 45.00 | 1.00 | 4.23E-04 |
| 2 | 0.7070 | 0.7071 | 432.78 | 0.56 | 45.00 | 1.00 | 1.29E-03 |
| 3 | 0.7070 | 0.7071 | 432.32 | 1.23 | 45.00 | 1.00 | 2.84E-03 |
| 4 | 0.7070 | 0.7071 | 431.30 | 2.14 | 45.00 | 1.00 | 4.96E-03 |
| 5 | 0.7071 | 0.7070 | 429.33 | 3.37 | 45.00 | 1.00 | 7.85E-03 |
| 7 | 0.7071 | 0.7071 | 426.44 | 6.54 | 45.00 | 1.00 | 1.53E-02 |
| 10 | 0.7071 | 0.7070 | 420.01 | 12.90 | 45.00 | 1.00 | 3.06E-02 |

In this case $I1$ and $I2$ are identical but $I2$ is rotated laterally by an amount of angle θ as shown in Figure 5.11. Figure 5.12 shows the data plot and the eigenvectors. As can be seen, the angle θ_1 remains 45° although the data points are spreading away from the eigenvector. Indeed changing the angle θ will only change the eigenvalues but not the directions of the eigenvectors. It should be note that the change of parameters is the same fashion as the shift, in case 2. Table 5.3 shows the resulting parameters as θ is changed.

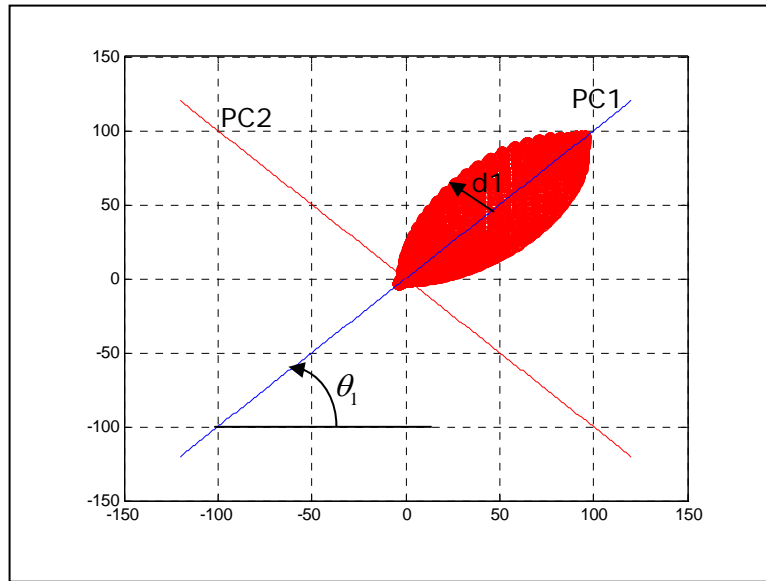


Figure 5.12 Data points on the PCA axis ; $\theta = 10^\circ$

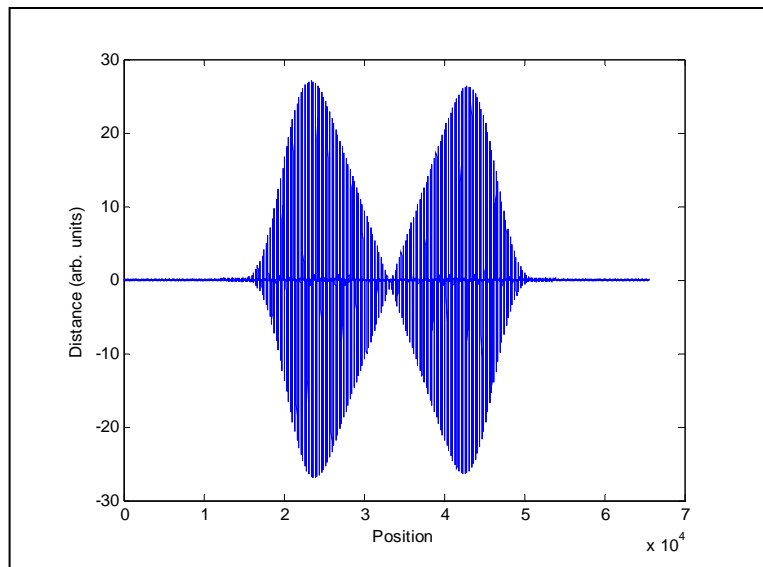


Figure 5.13 Graph of distance **d1** with $\theta = 10^\circ$

In Figure 5.13, we plot distance **d1**, see Figure 5.12, as a function of the locations of the data points. In contrast to figure 5.10, we now have a butterfly shape with a zero

at the centre and the positive and the negative parts being symmetrical about the eigenvector.

Comparing Table 5.2 to Table 5.3, both have the eigenvectors = 45 degrees, although the eigenvalues ratios change differently, it is not apparent what information one can derive to distinguish between the two. Thus to classify both cases further we will use the 1D PCA again and examine the profiles of image. The procedures are described as following.

Step 1: Take profiles of template and sample image

An image of the template and its profile taking along line P0, about the centre, are shown in Figure 5.14 and 5.15. For the sample image, several profiles, Ps0 to PsN, are taken with distance D from the centre profile Ps0, distance D is shown in Figure 5.16. Example of the sample images and associated profiles for each case are shown in Figure 5.16, 5.17, 5.18 and 5.19.

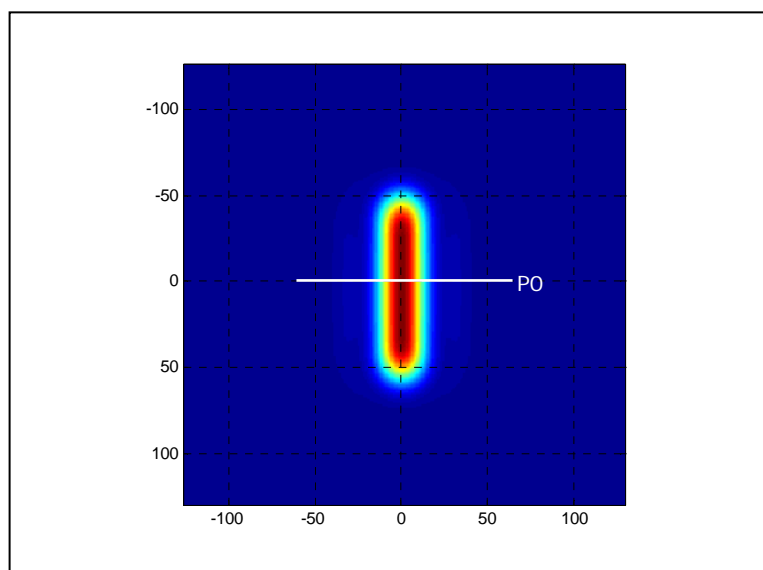


Figure 5.14 An example of template image

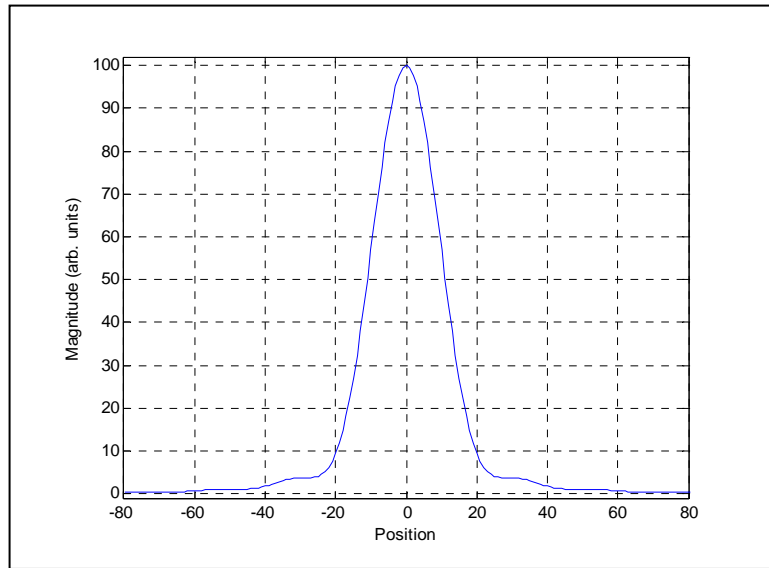


Figure 5.15 An example profile of template image takes along line P0

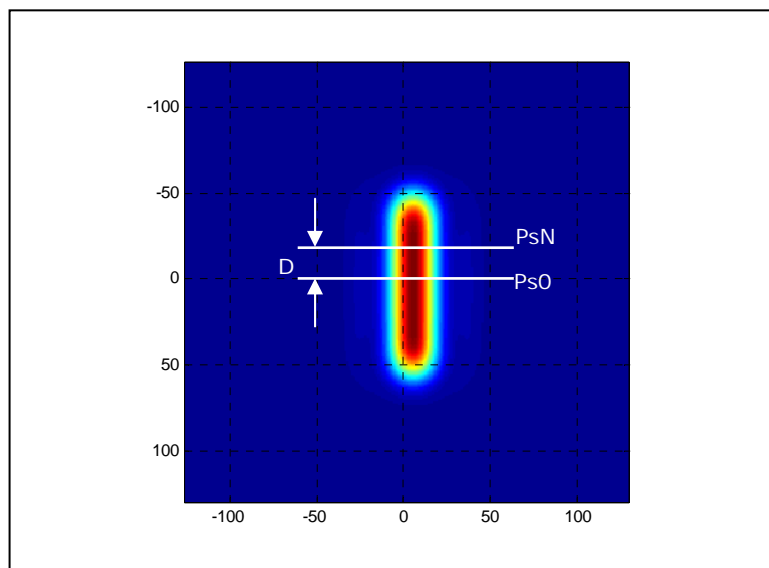


Figure 5.16 Shifted sample image

From Figure 5.16, the sample image is shifted by Δ . Profiles, Ps0 to PsN, are taken along the length of the image with different distance D. In this case every profiles have the same shifted Δ . An example of the profile is shown in Figure 5.17.

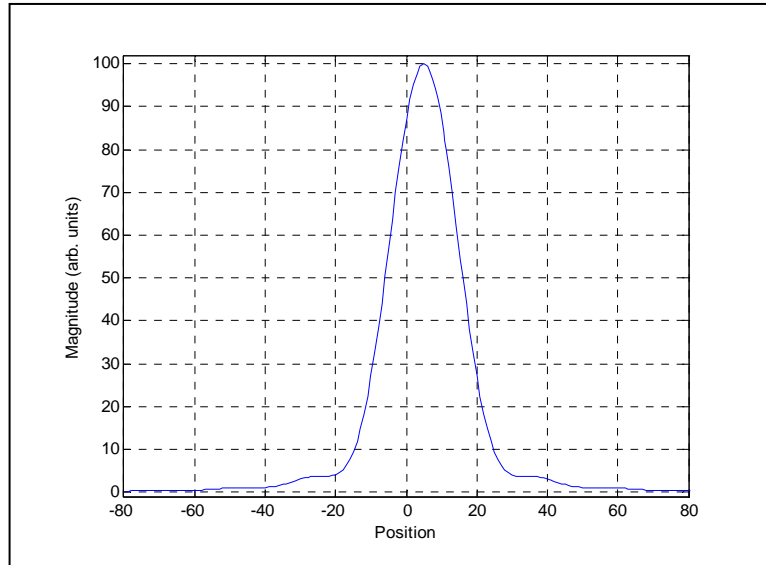


Figure 5.17 An example profile of shifted image

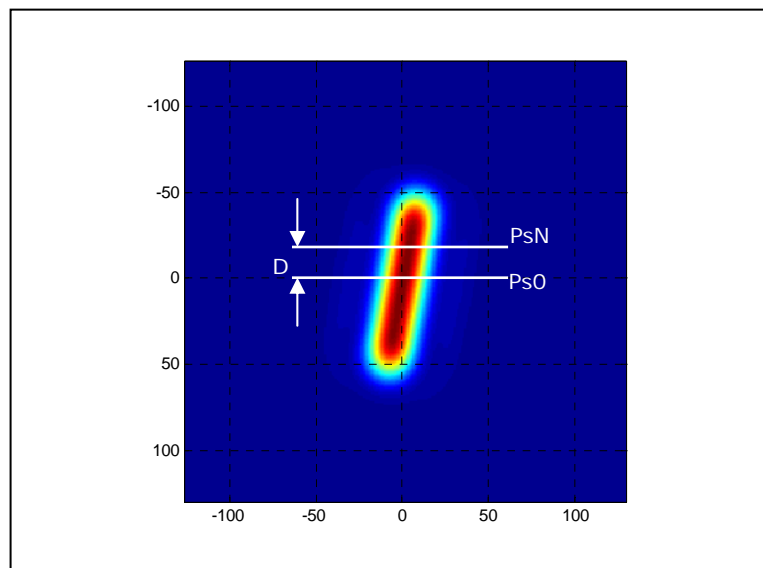


Figure 5.18 Rotated sample image

From Figure 5.18, the sample image is rotated by θ . Profiles are taken the same way as in the shift case. The graphs of the profiles are shown in Figure 5.19. As can be seen each profile is shifted from the centre according to the distance D .

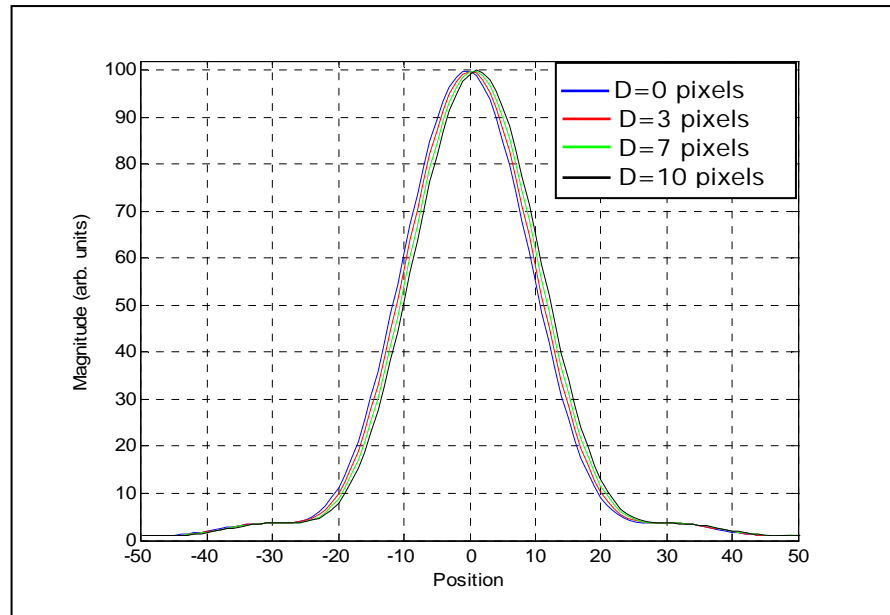


Figure 5.19 Example profiles of rotated image

Step 2: Perform PCA between profiles of template P_0 and samples

$$\text{PCA}\{P_0, P_{S_m}\}$$

5-1

Where $m=0$ to N .

2.1 In case sample is shifted

Performing PCA of the template P_0 and the sample P_{S_m} will yield the same parameters. Table 5.4 summarises the parameters resulted from performing PCA with shift $\Delta = 5$ pixels as the distance D of the profile increases. It should be noted that the eigenvalues ratio, the last column, are not zero although the eigenvectors remain 45 degrees. This means that there is a shift between the template and the sample. Thus we can use the shifting property

of FFT as have been described in Chapter 4, section 4.6.1, to shift the image accurately.

Table 5.4 Result of PCA

| Distance D pixels | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 3 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |
| 5 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |
| 7 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |
| 10 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |
| 13 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |
| 15 | 0.7071 | 0.7071 | 1073.59 | 45.38 | 45.00 | 1.00 | 4.23E-02 |

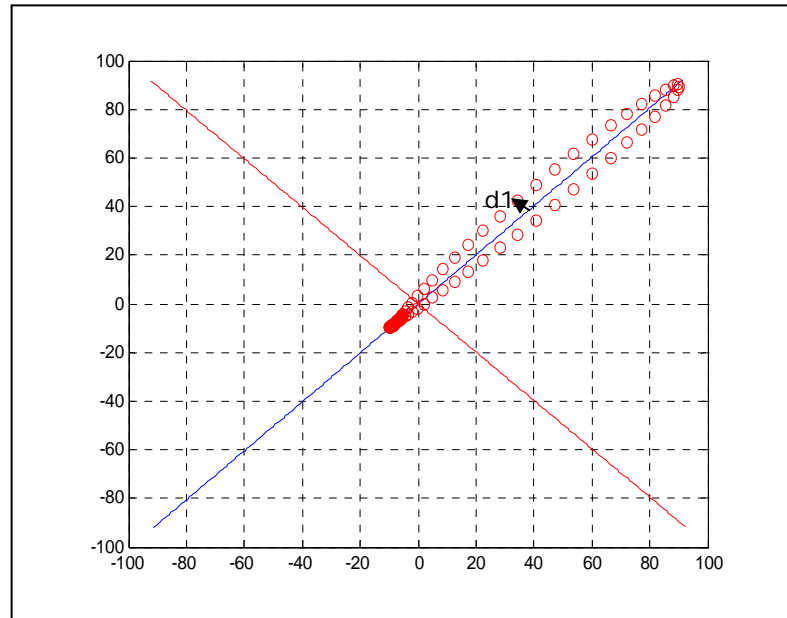
2.2 Sample is rotated by 5 degrees

Performing PCA on the profiles and tabulating the parameters result in Table 5.5. It can be seen that the eigenvalues ratio change according to the increase of profile distances D although the eigenvectors remain 45 degrees. This means the sample is rotated with some amount of θ .

Table 5.5 Result of PCA for 5 degrees rotation

| Distance D pixels | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|-------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 3 | 0.7071 | 0.7071 | 1120.66 | 0.0047 | 45.00 | 1.00 | 4.15E-06 |
| 5 | 0.7071 | 0.7071 | 1120.85 | 0.0876 | 45.00 | 1.00 | 7.81E-05 |
| 7 | 0.7071 | 0.7071 | 1121.36 | 0.2875 | 45.00 | 1.00 | 2.56E-04 |
| 10 | 0.7071 | 0.7071 | 1121.84 | 0.8033 | 45.00 | 1.00 | 7.17E-04 |
| 13 | 0.7071 | 0.7071 | 1119.57 | 1.5678 | 45.00 | 1.00 | 1.40E-03 |
| 15 | 0.7071 | 0.7071 | 1118.99 | 2.2112 | 45.00 | 1.00 | 1.98E-03 |

Figure 5.20 shows the data plot and the eigenvectors for $D = 9$ pixels. A 10 degrees rotation has been used in order to show clearly the spreading of the data points from the eigenvector. We represent distances of each data point from the eigenvector with **d1**, show in the Figure 5.20, and then plot **d1** as a function of its position in Figure 5.21.

Figure 5.20 Data points on the PCA axis ; $\theta = 10^\circ$

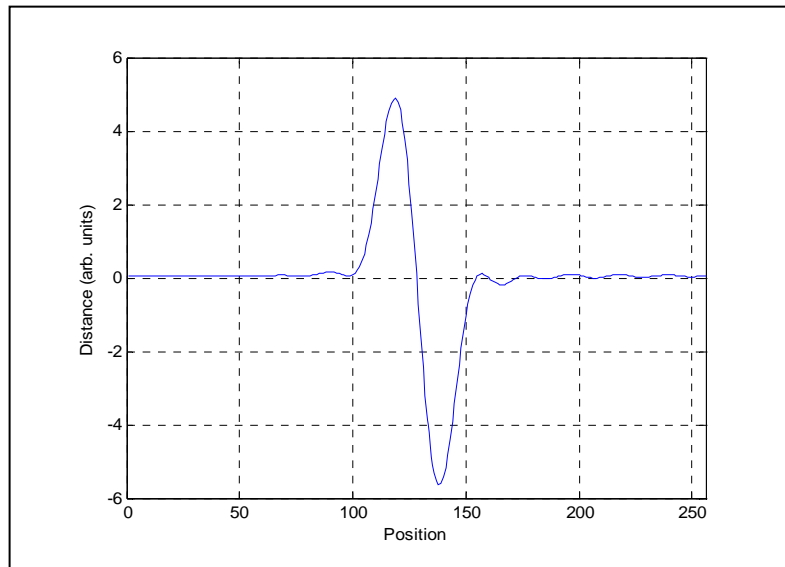


Figure 5.21 Graph of distance **d1** with $\theta = 10^\circ$

We calculate the standard deviation of **d1** for different angle θ and distances of profiles D. The results are tabulated in Table 5.6 and then plot graphs of $\text{STD}(\text{d1})$ as a function of distances D shown in Figure 5.22. It can be seen that each graph, different angle θ , increases linearly according to D. It should be noted that as long as the distances D are less than 40 percents of images length, the graphs remain linear.

Table 5.6 Standard deviation of **d1** with different angles and distances D

| Distance D (pixels) | Angle θ | | | | | | | | | |
|---------------------------|----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 1° | 2° | 3° | 4° | 5° | 6° | 7° | 8° | 9° | 10° |
| 0 | 0.0116 | 0.0235 | 0.0367 | 0.0518 | 0.0696 | 0.0905 | 0.1147 | 0.1425 | 0.1740 | 0.2093 |
| 1 | 0.0356 | 0.0708 | 0.1061 | 0.1417 | 0.1780 | 0.2153 | 0.2539 | 0.2941 | 0.3362 | 0.3804 |
| 3 | 0.0837 | 0.1670 | 0.2501 | 0.3331 | 0.4161 | 0.4994 | 0.5831 | 0.6673 | 0.7520 | 0.8374 |
| 5 | 0.1319 | 0.2633 | 0.3946 | 0.5257 | 0.6569 | 0.7878 | 0.9187 | 1.0497 | 1.1810 | 1.3127 |
| 7 | 0.1801 | 0.3594 | 0.5387 | 0.7178 | 0.8965 | 1.0749 | 1.2531 | 1.4314 | 1.6103 | 1.7899 |
| 9 | 0.2282 | 0.4549 | 0.6820 | 0.9083 | 1.1342 | 1.3597 | 1.5854 | 1.8118 | 2.0390 | 2.2651 |
| 11 | 0.2758 | 0.5498 | 0.8239 | 1.0973 | 1.3699 | 1.6427 | 1.9165 | 2.1899 | 2.4623 | 2.7349 |
| 13 | 0.3230 | 0.6442 | 0.9650 | 1.2849 | 1.6044 | 1.9251 | 2.2450 | 2.5637 | 2.8827 | 3.2030 |
| 15 | 0.3698 | 0.7381 | 1.1055 | 1.4718 | 1.8390 | 2.2060 | 2.5712 | 2.9364 | 3.3035 | 3.6683 |
| 17 | 0.4168 | 0.8321 | 1.2461 | 1.6595 | 2.0744 | 2.4864 | 2.8980 | 3.3115 | 3.7221 | 4.1317 |
| 19 | 0.4644 | 0.9269 | 1.3879 | 1.8493 | 2.3101 | 2.7686 | 3.2284 | 3.6864 | 4.1421 | 4.6001 |
| 21 | 0.5133 | 1.0233 | 1.5317 | 2.0419 | 2.5485 | 3.0547 | 3.5616 | 4.0640 | 4.5683 | 5.0690 |
| 23 | 0.5640 | 1.1214 | 1.6781 | 2.2358 | 2.7899 | 3.3458 | 3.8966 | 4.4472 | 4.9964 | 5.5411 |
| 25 | 0.6162 | 1.2210 | 1.8266 | 2.4319 | 3.0345 | 3.6372 | 4.2348 | 4.8340 | 5.4251 | 6.0193 |

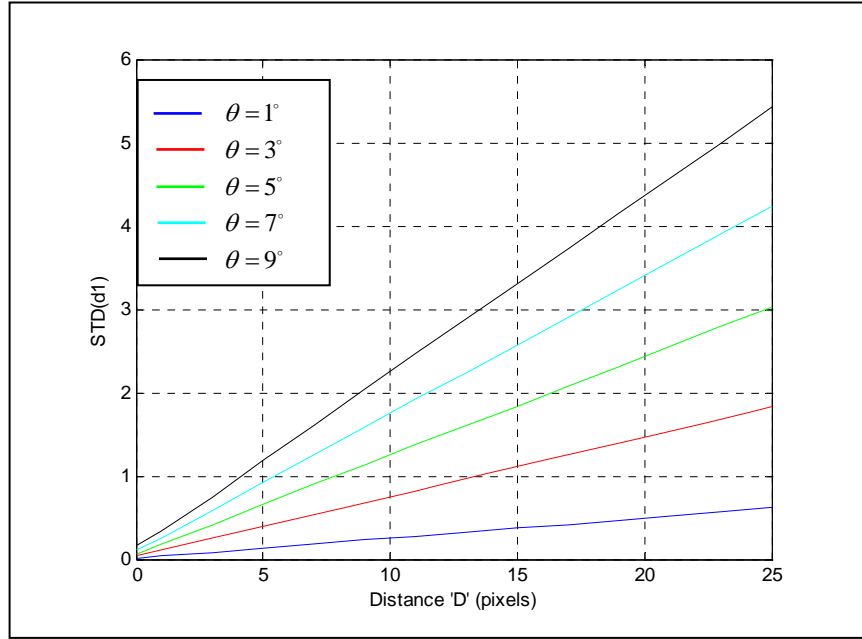


Figure 5.22 Graph of standard deviation of **d1** with different θ

Base on the Table 5.6 or graphs in Figure 5.22 where we can select the distance D to perform PCA and then calculate standard deviation of **d1** to determine a rotation angle of the sample. Repeating this process with the aim of minimising the error signal (which can be defined as the absolute sum of the area of the graph in figure 5.21, or its standard deviation), accurate rotation of the original image can be achieved.

Case 4: $I2$ is rotated by θ and shifted by Δ

The line segments $I1$ and $I2$ are identical, but $I2$ is rotated by an angle $\theta = 5^\circ$ and shifted by $\Delta y = 5$ pixels. The sizes of the line segment used in the simulation are length = 101 pixels and width = 13 pixels. Table 5.7 is tabulated with $\Delta x = 0$ to 6 pixels. From the result it can be seen the angle θ_1 of the principal vector remains

45° although $I2$ has been shifted and rotated. Thus at this point we can conclude that changing the position and orientation of $I2$ will only change the eigenvalues but not the directions of the eigenvectors. This means we can determine the size of $I2$, $I1=I2$, by only checking the angle θ_1 of the eigenvector being 45° .

Table 5.7 Result of PCA

| $I2$ shifted Δx (pixels) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|--|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0 | 0.7071 | 0.7070 | 427.29 | 5.40 | 45.00 | 1.00 | 1.26E-02 |
| 1 | 0.7071 | 0.7070 | 426.95 | 5.75 | 45.00 | 1.00 | 1.35E-02 |
| 2 | 0.7071 | 0.7070 | 425.35 | 7.35 | 45.00 | 1.00 | 1.73E-02 |
| 3 | 0.7071 | 0.7070 | 422.51 | 10.18 | 45.00 | 1.00 | 2.41E-02 |
| 4 | 0.7071 | 0.7070 | 418.50 | 14.20 | 45.00 | 1.00 | 3.39E-02 |
| 5 | 0.7071 | 0.7070 | 413.37 | 19.32 | 45.00 | 1.00 | 4.67E-02 |
| 6 | 0.7071 | 0.7070 | 407.21 | 25.48 | 45.00 | 1.00 | 6.26E-02 |

To determine the correct position and orientation of the sample images in this case, the orientation will be determined first and then the position by using the processes described in **case 3**.

Case 5: $I1$ and $I2$ are different in sizes (Width and/or Length)

Firstly we will examine the case where the lengths of $I1$ and $I2$ are the same. $I2$ will have different values in width whilst that of $I1$ is constant. The lengths of $I1$ and $I2$ are 101 pixels and the width of $I1$ is 13 pixels. Table 5.8 shows the resulting parameters when performing PCA of $I1$ and $I2$. As can be seen, if the width of $I2$ is

smaller than $I1$, the angle θ_1 is smaller than 45° . In contrast when the width of $I2$ is greater than $I1$, the angle θ_1 becomes bigger than 45° . Another point is that the ratio of eigenvalues $\Lambda_{11} / \Lambda_{22}$ is non-zero according to the differences of the width of $I1$ and $I2$.

Table 5.8 Result of PCA

| Object (I2) width (pixels) | Object widths Ratio | Profile Widths image Ratio | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12} / V_{22} | $\Lambda_{11} / \Lambda_{22}$ |
|-------------------------------------|---------------------------|-------------------------------------|----------|----------|----------------|----------------|------------|-------------------|-------------------------------|
| 3 | 0.23 | 0.92 | 0.7196 | 0.6944 | 417.73 | 0.44 | 43.98 | 1.04 | 1.06E-03 |
| 5 | 0.38 | 0.93 | 0.7184 | 0.6957 | 419.22 | 0.36 | 44.08 | 1.03 | 8.57E-04 |
| 7 | 0.54 | 0.94 | 0.7165 | 0.6975 | 421.46 | 0.25 | 44.23 | 1.03 | 5.90E-04 |
| 9 | 0.69 | 0.95 | 0.7141 | 0.7001 | 424.49 | 0.13 | 44.43 | 1.02 | 3.14E-04 |
| 11 | 0.85 | 0.97 | 0.7109 | 0.7033 | 428.33 | 0.04 | 44.69 | 1.01 | 9.20E-05 |
| 13 | 1 | 1 | 0.7071 | 0.7071 | 433.02 | 0.00 | 45.00 | 1.00 | 0 |
| 15 | 1.15 | 1.03 | 0.7026 | 0.7116 | 438.60 | 0.05 | 45.37 | 0.99 | 1.20E-04 |
| 17 | 1.31 | 1.06 | 0.6972 | 0.7168 | 445.11 | 0.24 | 45.79 | 0.97 | 5.38E-04 |
| 19 | 1.46 | 1.10 | 0.6911 | 0.7227 | 452.60 | 0.61 | 46.28 | 0.96 | 1.34E-03 |
| 21 | 1.62 | 1.14 | 0.6842 | 0.7293 | 461.12 | 1.20 | 46.83 | 0.94 | 2.59E-03 |
| 23 | 1.77 | 1.19 | 0.6765 | 0.7365 | 470.70 | 2.06 | 47.43 | 0.92 | 4.37E-03 |

Note: 1 pixel = 38.6 nm

In the case when the widths of $I1$ and $I2$ are equal, the length of $I2$ is changed whilst $I1$ is constant. The widths of $I1$ and $I2$ are both 13 pixels and the length of $I1$ is 101 pixels. Table 5.9 shows the resulting parameters when performing PCA of $I1$ and $I2$.

From Table 5.9, it can be seen the change of parameters are the same fashion as in the case varying $I2$ width.

Table 5.9 Result of PCA

| Object (I2) Length (pixels) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|--------------------------------------|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 91 | 0.7262 | 0.6875 | 409.08 | 1.67 | 43.43 | 1.06 | 4.08E-03 |
| 93 | 0.7221 | 0.6918 | 414.22 | 1.08 | 43.77 | 1.04 | 2.61E-03 |
| 95 | 0.7182 | 0.6958 | 419.15 | 0.61 | 44.09 | 1.03 | 1.46E-03 |
| 97 | 0.7144 | 0.6997 | 423.90 | 0.27 | 44.40 | 1.02 | 6.45E-04 |
| 99 | 0.7108 | 0.7034 | 428.52 | 0.07 | 44.70 | 1.01 | 1.60E-04 |
| 101 | 0.7071 | 0.7071 | 433.02 | 0.00 | 45.00 | 1.00 | 0 |
| 103 | 0.7035 | 0.7107 | 437.39 | 0.07 | 45.29 | 0.99 | 1.58E-04 |
| 105 | 0.6999 | 0.7142 | 441.65 | 0.28 | 45.58 | 0.98 | 6.24E-04 |
| 107 | 0.6964 | 0.7177 | 445.79 | 0.62 | 45.86 | 0.97 | 1.39E-03 |
| 109 | 0.6929 | 0.7211 | 449.81 | 1.10 | 46.14 | 0.96 | 2.44E-03 |
| 110 | 0.6894 | 0.7243 | 453.60 | 1.71 | 46.41 | 0.95 | 3.76E-03 |

We can conclude the result of performing PCA in term of angle θ_1 in two cases, **T1** and **T2**. We will use flow charts to demonstrate, when the angle is not 45 degrees, the procedures of selecting different templates for the purpose of PCA. It should be emphasised that the initial selection of a new template is not important, apart from the fact that it is easier to program the algorithm if we select a different length first.

For convenience, we use W1 and L1 to represent the width and length of $I1$ and W2 and L2 for $I2$, respectively.

Case T1: $\theta_1 > 45^\circ$

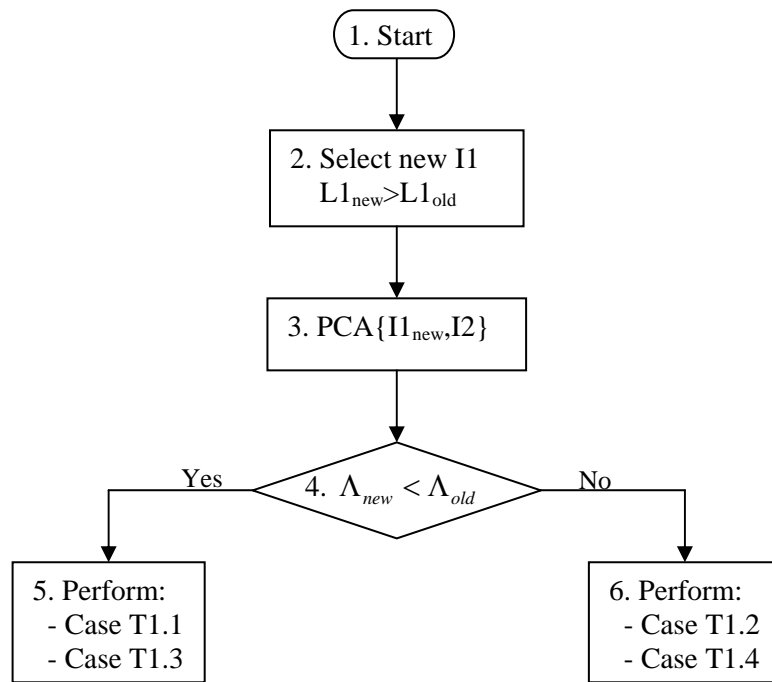
T1.1 ($W1=W2$) and ($L1<L2$)

T1.2 ($W1<W2$) and ($L1=L2$)

T1.3 ($W1<W2$) and ($L1<L2$)

T1.4 ($W1<W2$) and ($L1>L2$)

The following flow charts show the procedure to select new template for this case.



Flow chart 5.1 Procedure for selecting new template in case $\theta_1 > 45^\circ$

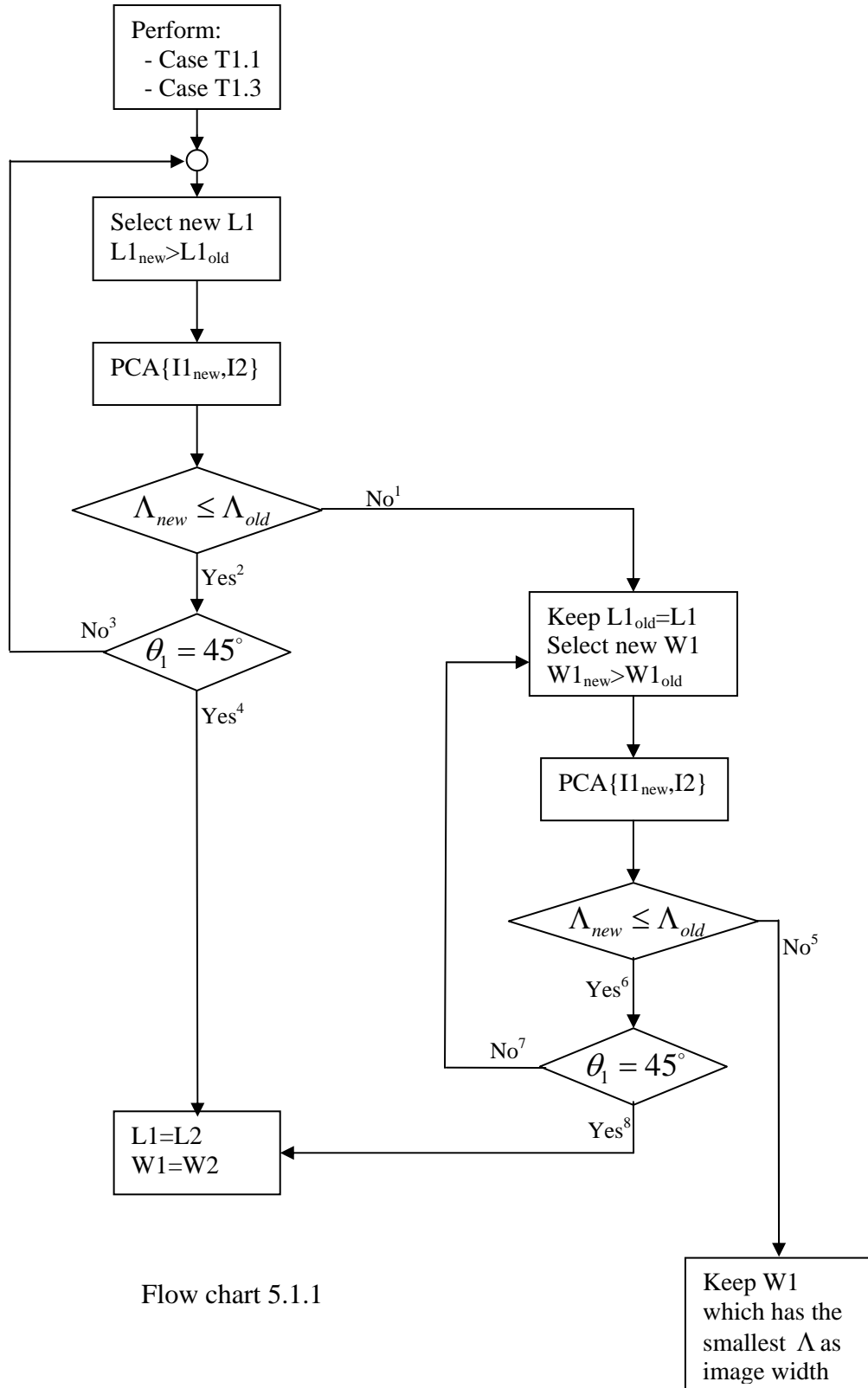
The meaning of each box in the Flow chart 5.1 can be explained as following.

1. After an initial PCA, with result $\theta > 45^\circ$
2. Select new **I1** of longer length
3. Perform PCA{**I1**_{new}, **I2**}
4. Test eigenvalues ratio, Λ_{new} and Λ_{old} then goto 5 or 6
 - Yes : length of **I1** is shorter than **I2**

- No : length of **I1** is longer than **I2**

5. or 6. Perform PCA of individual cases with new templates

The procedure in boxes 5 and 6 can be described by Flow chart 5.1.1 and 5.1.2 respectively.



The meaning of each superscript in the Flow chart 5.1.1 is described as following.

¹The length of Image **I1** is longer than **I2**; keep the old length of **I1**

² A new length of Image **I1** is shorter than **I2** or the length of **I1=I2**

³ Size of Image **I1** and **I2** is different

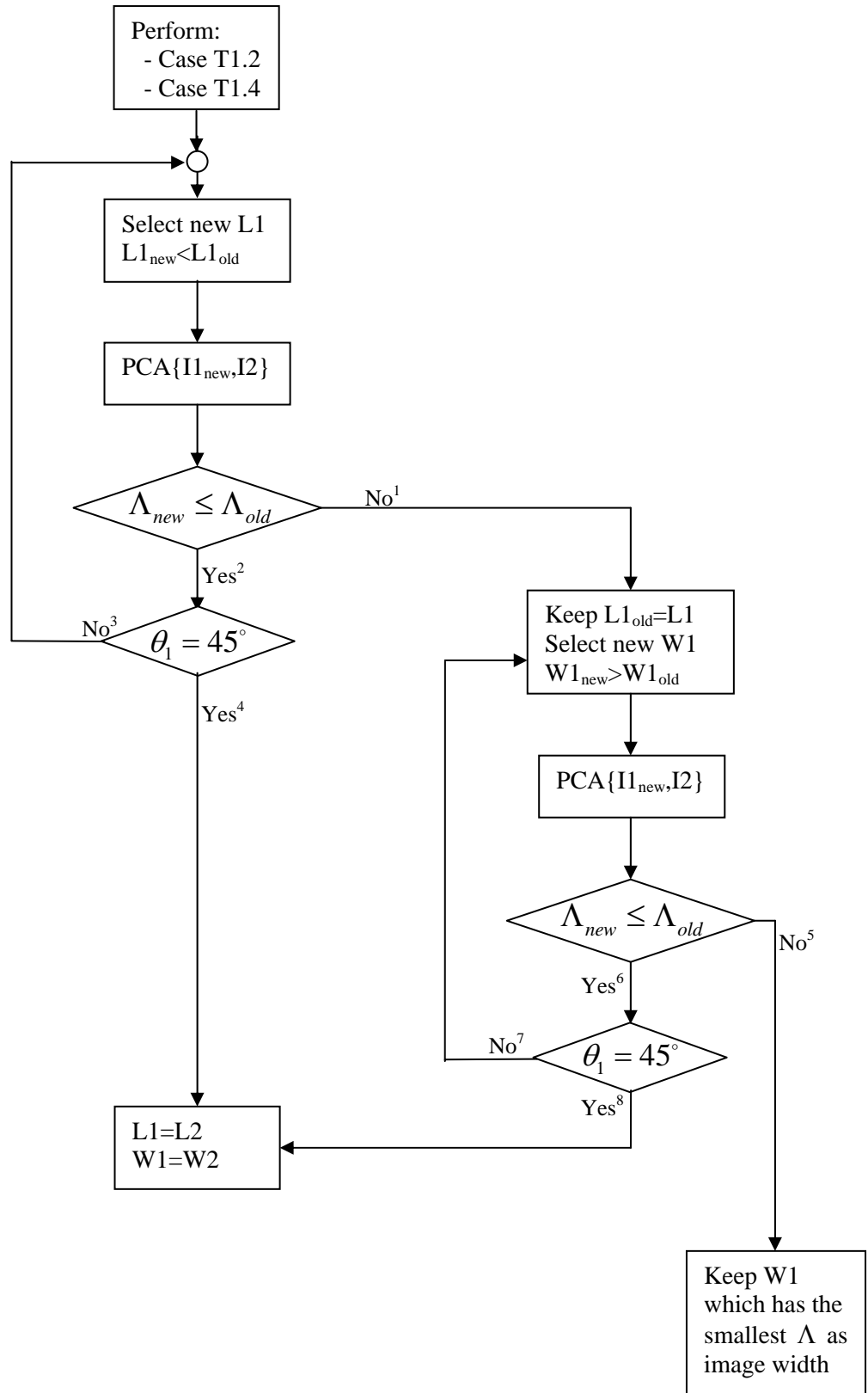
⁴ Size of Image **I1** and **I2** is match (**I1=I2**)

⁵ Image **I1** is wider than **I2**; keep the old width of **I1**

⁶ A new width of **I1** is smaller than **I2**

⁷ Size of Image **I1** and **I2** is different

⁸ Size of Image **I1** and **I2** is match (**I1=I2**)



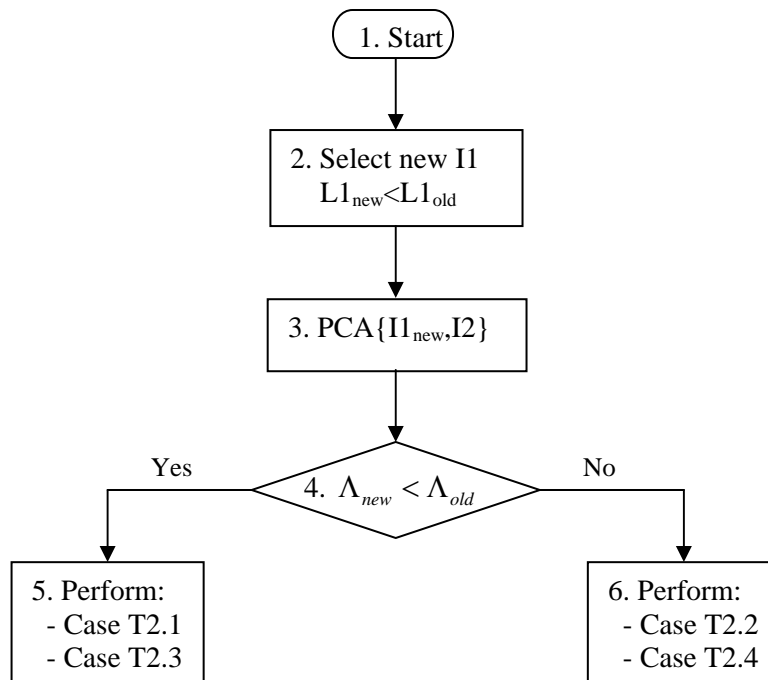
Flow chart 5.1.2

The meaning of each superscript in the Flow chart 5.1.2 is described as following.

- ¹The length Image **I1** is shorter than **I2**; keep the old length of **I1**
- ² A new length of Image **I1** is longer than **I2** or the length of **I1=I2**
- ³ Size of Image **I1** and **I2** is different
- ⁴ Size of Image **I1** and **I2** is match (**I1=I2**)
- ⁵ Image **I1** is wider than **I2**; keep the old width of **I1**
- ⁶ A new width of **I1** is smaller than **I2** or the width of **I1=I2**
- ⁷ Size of Image **I1** and **I2** is different
- ⁸ Size of Image **I1** and **I2** is match (**I1=I2**)

Case T2: $\theta_1 < 45^\circ$ T2.1 ($W1=W2$) and ($L1>L2$)T2.2 ($W1>W2$) and ($L1=L2$)T2.3 ($W1>W2$) and ($L1>L2$)T2.4 ($W1>W2$) and ($L1<L2$)

The following flow charts show the procedure to select new template for this case.



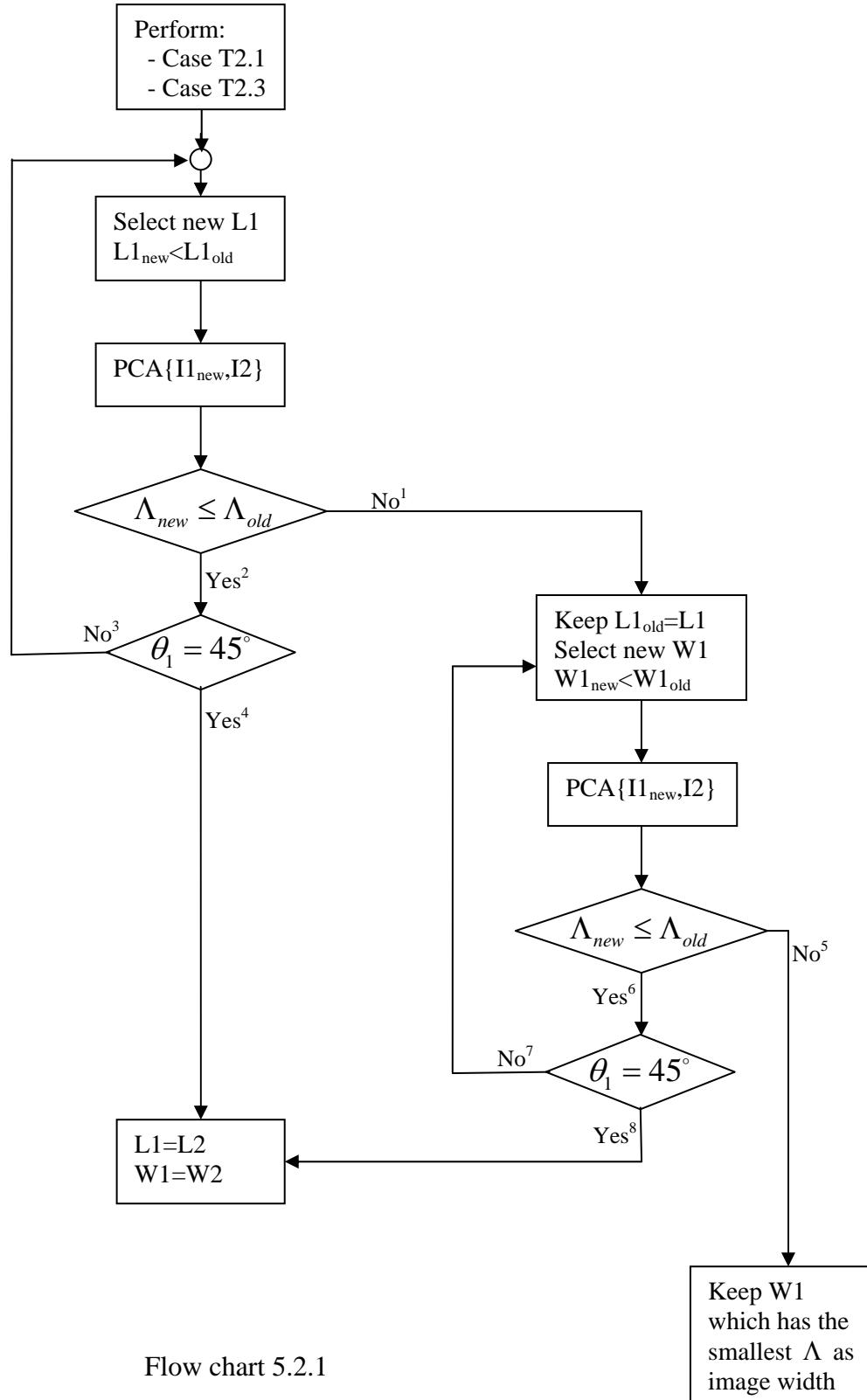
Flow chart 5.2

The meaning of each box in the Flow chart 5.2 can be explained as following.

1. After an initial PCA, with result $\theta < 45^\circ$
2. Select new *I1* of shorter length
3. Perform PCA{*I1*_{new}, *I2*}
4. Test eigenvalues ratio, Λ_{new} and Λ_{old} then goto 5 or 6
 - Yes : length of **I1** is longer than **I2**
 - No : length of **I1** is shorter than **I2**

5. or 6. Perform PCA of individual cases with new templates

The procedure in boxes 5 and 6 can be described by Flow chart 5.2.1 and 5.2.2 respectively.



The meaning of each superscript in the Flow chart 5.2.1 is described as following.

¹The length Image **I1** is shorter than **I2**; keep the old length of **I1**

² A new length of Image **I1** is longer than **I2** or the length of **I1=I2**

³ Size of Image **I1** and **I2** is different

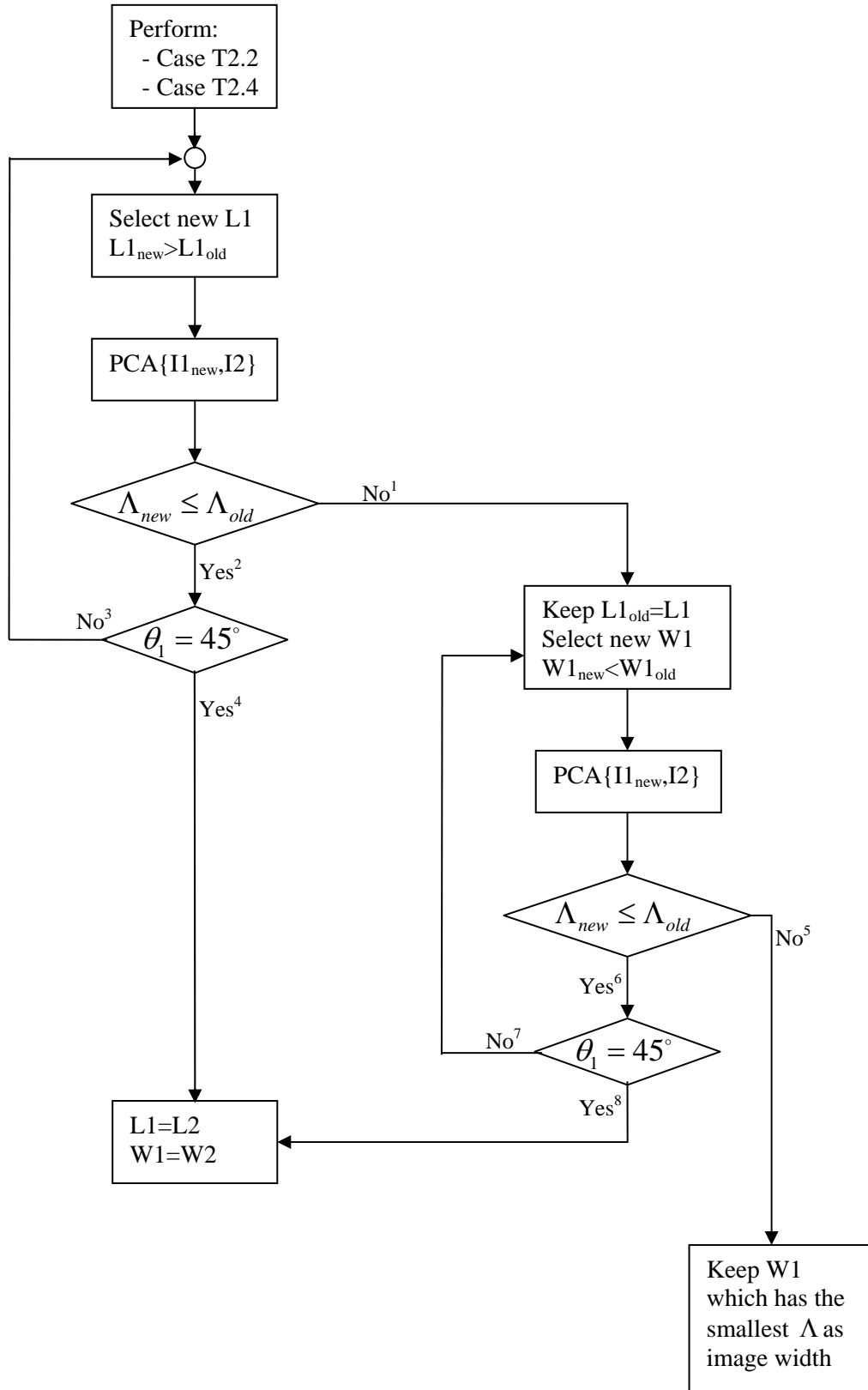
⁴ Size of Image **I1** and **I2** is match (**I1=I2**)

⁵ The width of image **I1** is smaller than **I2**; keep the old width of **I1**

⁶ A new width of **I1** is wider than **I2** or the width of **I1=I2**

⁷ Size of Image **I1** and **I2** is different

⁸ Size of Image **I1** and **I2** is match (**I1=I2**)



Flow chart 5.2.2

The meaning of each superscript in the Flow chart 5.2.2 is described as following.

¹The length Image **I1** is longer than **I2**; keep the old length of **I1**

² A new length of Image **I1** is shorter than **I2** or the length of **I1=I2**

³ Size of Image **I1** and **I2** is different

⁴ Size of Image **I1** and **I2** is match (**I1=I2**)

⁵ The width of image **I1** is smaller than **I2**; keep the old width of **I1**

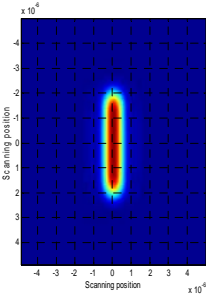
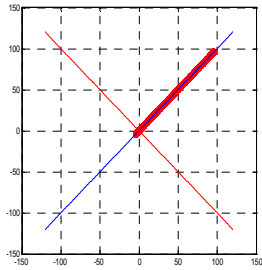
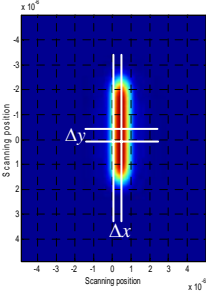
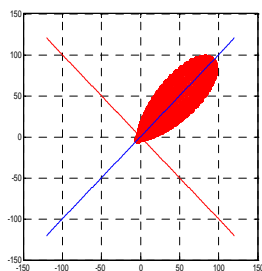
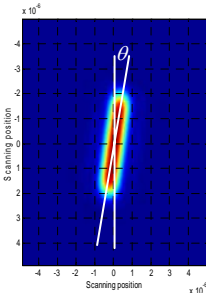
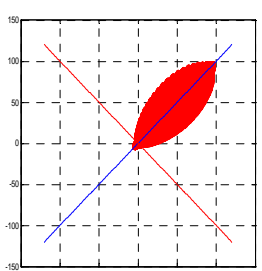
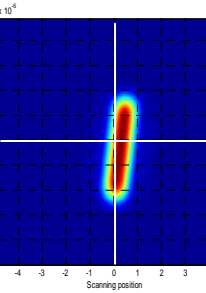
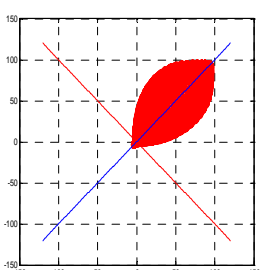
⁶ A new width of **I1** is wider than **I2** or the width of **I1=I2**

⁷ Size of Image **I1** and **I2** is different

⁸ Size of Image **I1** and **I2** is match (**I1=I2**)

The distribution of data points plot on PCA axis for each case and its characteristics are summarised in Table 5.10.

Table 5.10 Characteristics and distributions of data points on PCA axis

| Case/Characteristics | Line segment images – I_2 | Distribution of data points on PCA axis |
|--|---|---|
| $- I_1 = I_2$ $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} = 0$ |  |  |
| $- I_2$ shifted by Δ_x and/or Δ_y $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} \neq 0$ |  |  |
| $- I_2$ rotated by angle θ $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} \neq 0$ |  |  |
| $- I_2$ rotated by angle θ and shifted by Δ_x and/or Δ_y $\theta_1 = 45^\circ$ $\Lambda_{11}/\Lambda_{22} \neq 0$ |  |  |

5.3 Non-straight line images

In this section we will demonstrate the capability of the technique to other shapes of objects. An example of the image is shown in Figure 5.23.

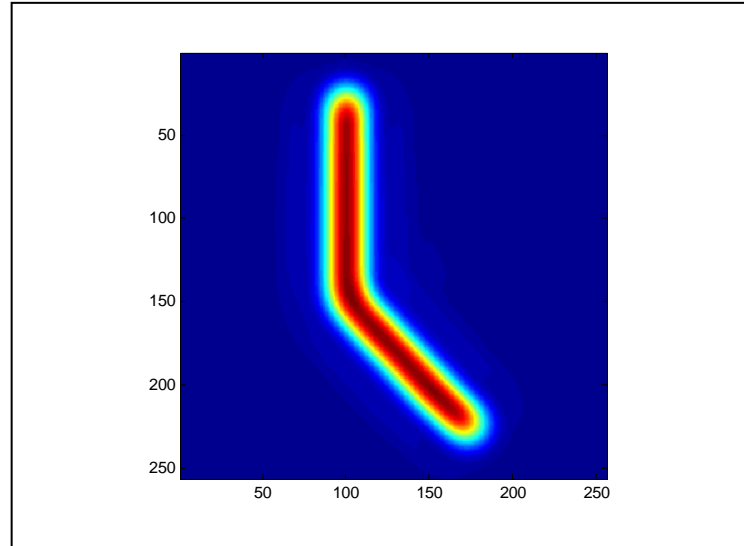
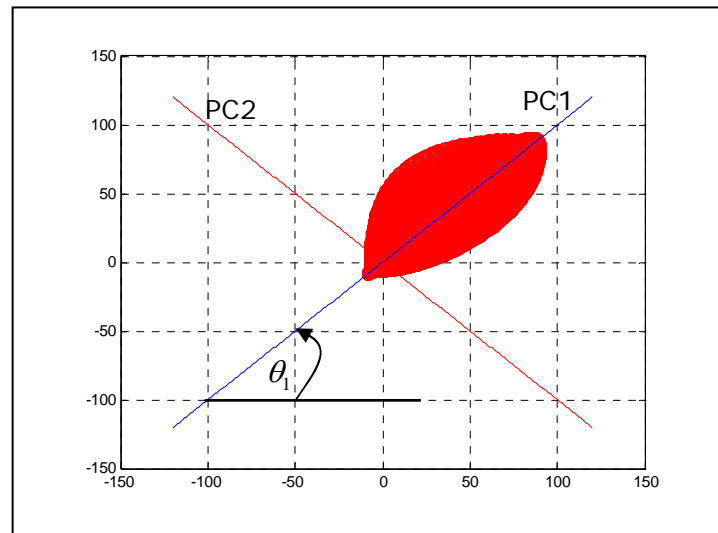


Figure 5.23 Non-straight line image

The images $I1$ and $I2$ are identical as shown in Figure 5.23, but with $I2$ rotated by an angle $\theta = 5^\circ$ and shifted by $\Delta y = 5$ pixels. Table 5.11 is tabulated the result of performing PCA with $\Delta x = 0$ to 10 pixels. From the result, it can be seen the angle θ_1 of the principal vector remains 45° although $I2$ is shifted and rotated. Thus at this point we can conclude that the shape of image does not affect the eigenvectors. This means we can determine the size of $I2$, $I1=I2$, by only checking the angle θ_1 of the eigenvector being 45° . Figure 5.24 shows the data plot and the eigenvectors. As can be seen, the spread of data points are not symmetrical about the eigenvector. For this reason, it is because of the shape of images.

Table 5.11 Result of PCA

| $I2$ shifted Δx (pixels) | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|--|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0 | 0.7071 | 0.7071 | 920.26 | 17.68 | 45.00 | 1.00 | 1.92E-02 |
| 1 | 0.7071 | 0.7071 | 919.85 | 18.88 | 45.00 | 1.00 | 2.05E-02 |
| 2 | 0.7071 | 0.7071 | 918.35 | 19.38 | 45.00 | 1.00 | 2.11E-02 |
| 3 | 0.7071 | 0.7071 | 917.71 | 20.01 | 45.00 | 1.00 | 2.18E-02 |
| 4 | 0.7071 | 0.7071 | 914.92 | 22.80 | 45.00 | 1.00 | 2.49E-02 |
| 5 | 0.7071 | 0.7071 | 910.00 | 27.71 | 45.00 | 1.00 | 3.05E-02 |
| 6 | 0.7071 | 0.7071 | 903.00 | 34.71 | 45.00 | 1.00 | 3.84E-02 |
| 7 | 0.7071 | 0.7071 | 894.00 | 43.70 | 45.00 | 1.00 | 4.89E-02 |
| 8 | 0.7071 | 0.7071 | 883.11 | 54.59 | 45.00 | 1.00 | 6.18E-02 |
| 9 | 0.7071 | 0.7071 | 870.46 | 67.24 | 45.00 | 1.00 | 7.72E-02 |
| 10 | 0.7071 | 0.7071 | 856.20 | 81.49 | 45.00 | 1.00 | 9.52E-02 |

Figure 5.24 Data points on the PCA axis ; $\theta = 5^\circ$, $\Delta x = 5$ pixels and $\Delta y = 5$ pixels

5.4 Coarse Measurement using Hough Transform

This section describes the use of Hough Transform to provide initial information of object for PCA. This information relate to location, orientation, shape and rough dimensions. It should be noted that this information only provide the initial parameters for PCA to select a proper template. For convenience, the image of an object in Figure 5.25 is used although the technique can be extended to other shape of objects.

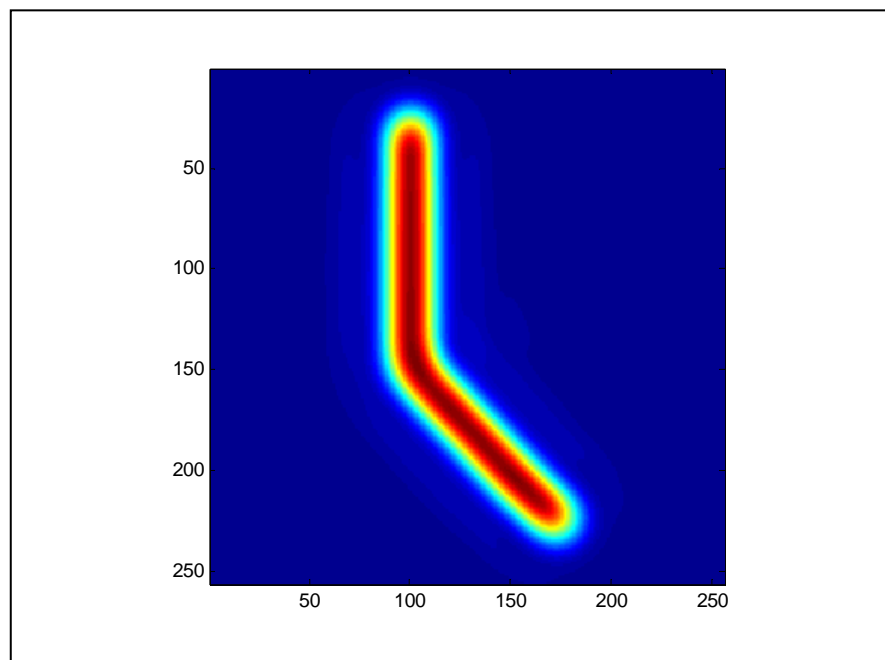


Figure 5.25 Example image of object

The Hough Transform space of the image shows in Figure 5.26 and 5.27. Figure 5.26 represents the accumulator $Acc1$ of HT space. The values of each array pixel in the space are the sum of magnitudes of image pixels of the same distance and angle.

Figure 5.27 is the accumulator $Acc2$ of HT space. The values of array pixel in this accumulator are the sum of number of image pixels of the same distance and angle. After that the maximum areas in the $Acc1$ can be determined (the details have been described in chapter 4, section 4.3). The detected maximum areas A1 and A2 are shown in Figure 5.28.

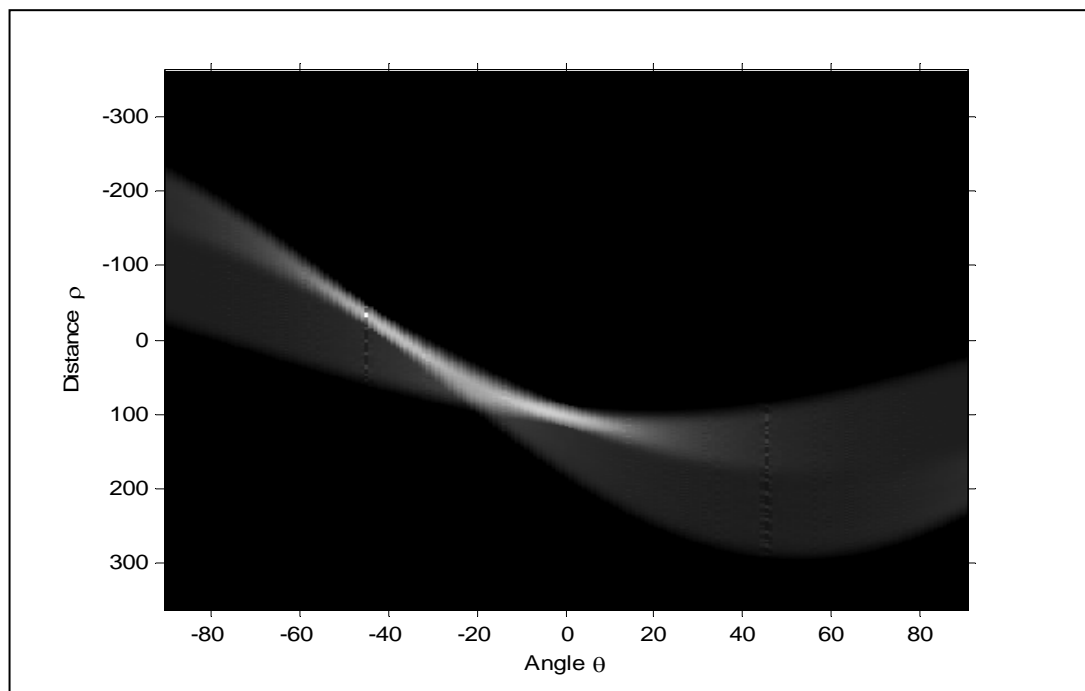


Figure 5.26 $Acc1$ of HT space

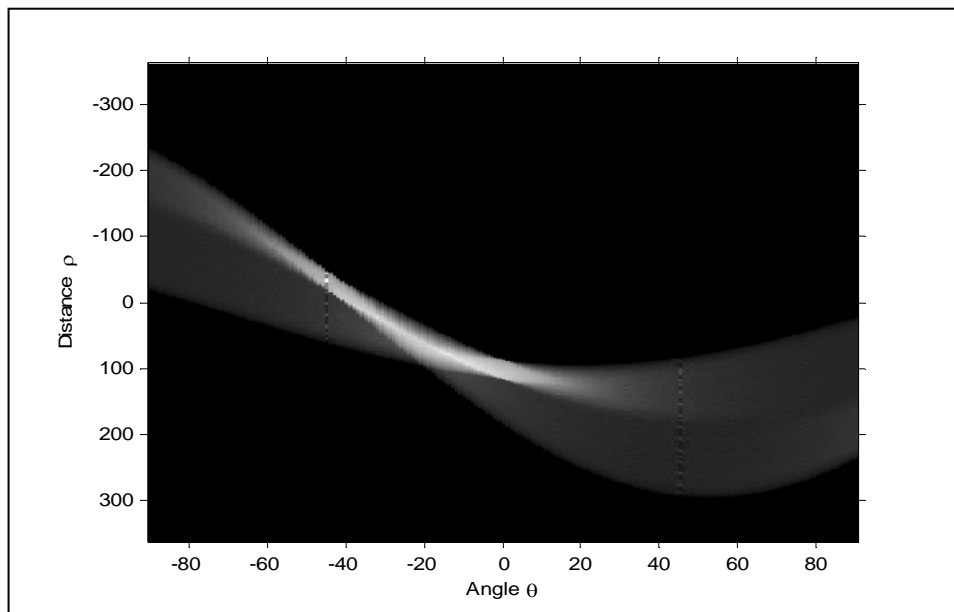


Figure 5.27 *Acc2* of HT space

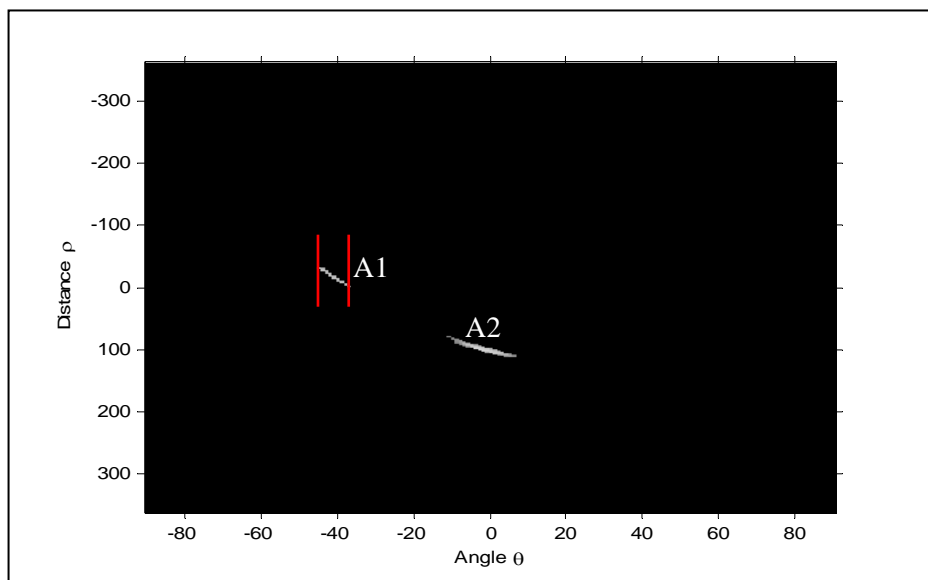


Figure 5.28 Maximum Areas from *Acc1*

Each area, A1 and A2, in Figure 5.28 represents a line segment in the figure. We can analyse individual area to determine parameters of the line segment such as the angle θ , distance ρ , width and length. Considering the area A1 first, we can determine an angular range of -38° to -44° , between the two red lines in Figure 5.28. Once the angles are established, the distances ρ of the line for each angle can be determined. The distance for each line segment is obtained by taking the location of the maximum point of the profile for each angle. Figure 5.29 shows the profiles taken from Figure 5.26 at these angles within the range -38° to -44° , and are plotted as a function of distance ρ . For example, considering the dotted blue graph in Figure 5.29, angle is -41° , the peak is determined to locate at $\rho = 18$ pixels, as indicated by the dashed line. The negative sign of distance ρ indicates that it is measured from the origin perpendicular to a line with negative angle. As has been discussed in Chapter 4, more precise measurement can be obtained by performing curve fitting near the peaks of the profiles. The next step is to determine the width and the length of the line segment. The widths and lengths of the line can be obtained by measuring the widths and heights of the profiles. The profile is taken at the same angle as used to determine the distance but this time taken from the HT space *Acc2*. Graphs of profile for each angle are shown in Figure 5.30. From Figure 5.30, consider the dotted blue graph which is associated with the same line segment as the dot blue graph in Figure 5.29. The width **W** of the line can be determined by measuring the width of the profile at 50% of its maximum magnitude. For the length **L** of the line segment is the mean value of the profile magnitude greater than 50% of its maximum magnitude. Thus we obtain the width of 35 pixels and length of 140 pixels. For the remaining profiles, each of them will have its own parameters. For area A2, we can analyse in the same way as for area A1. For example, one of the profile gives

parameters of a line segment image, angle $\theta = 0^\circ$, distance $\rho = 100$ pixels, width $W=35$ pixels and length $L=157$ pixels.

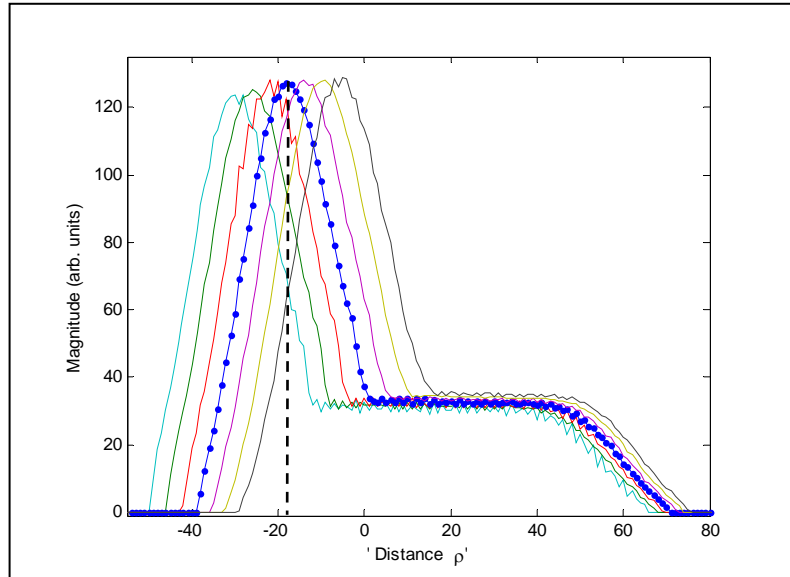


Figure 5.29 Profiles are taken from -38° to -44° in *Acc1*

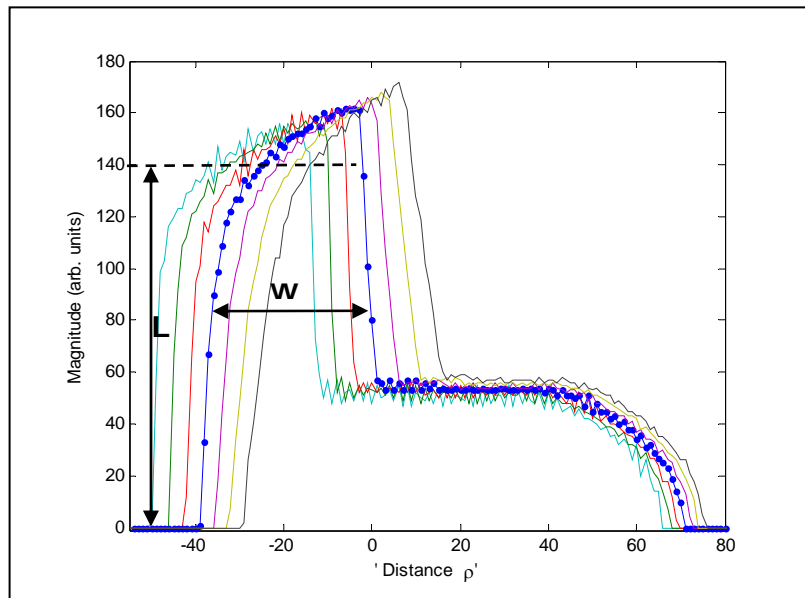


Figure 5.30 Profiles are taken from -38° to -44° from *Acc2*

Based on the parameters, θ , ρ and \mathbf{L} of each line image segment and by tracing back to the original image with the distance ρ as the middle of the line image segment, the skeletal line for each line image segment is obtained as shown in Figure 5.31. From Figure 5.31, it should be noted that the origin of the image plane is on the top left. With these skeletal lines the intersection \mathbf{P} of the lines can be determined, if it exists. Thus the shape of image is known. As the width \mathbf{W} of each line is known, we can widen the line according to its width. The roughly whole image is therefore reconstructed as shown in Figure 5.32. Thus the template for PCA should be chosen properly by using this image as the initial values.

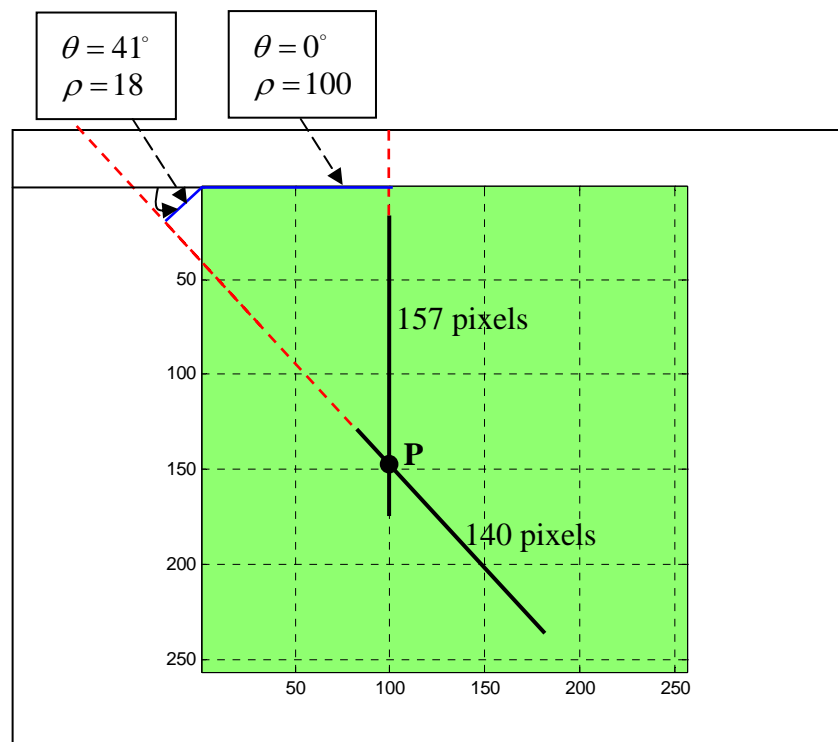


Figure 5.31 Detected skeletal line segments (black lines)

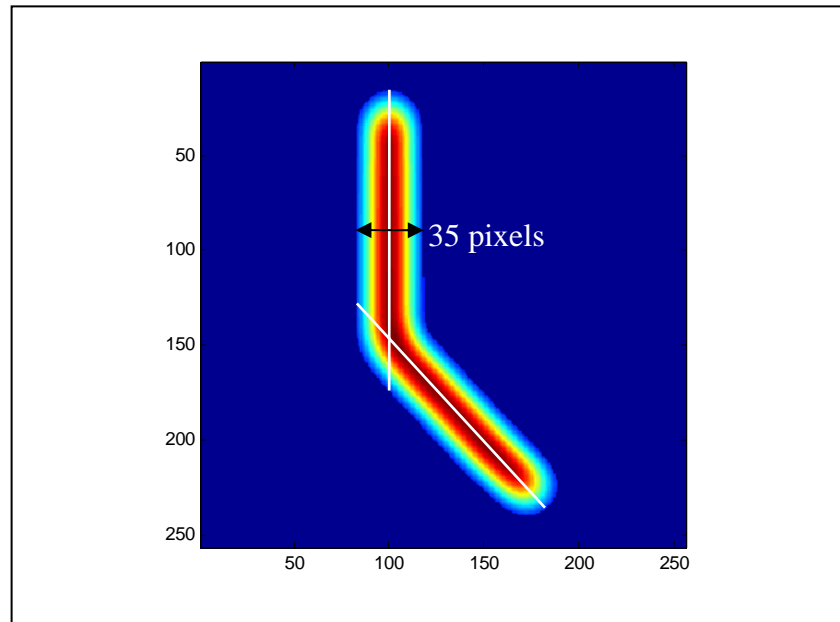


Figure 5.32 Determined whole image of object

The parameters of each line segment image obtained from area A1 or A2 in Figure 5.28. It also can be used as the initial information for PCA to measure individual line segment image without combining to each other.

Chapter 6

6 Experimental Results

Our technique has been illustrated using simulated data described in the previous chapter. In this chapter, the experimental results will be presented. The detailed step-by-step example of the processing steps is given. Firstly, the results from the Hough Transform are presented. This will be followed by result obtained from PCA.

6.1 Sample

The sample used is the resolving power test target USAF-1951[72]. The details of the sample are:

- Group 4/1 - 16 lines/pairs per millimetre
- Period = $62.5 \mu m$
- Bar width = $31.25 \mu m$
- Bar length = $156.25 \mu m$

An image of sample was taken by optical microscope with a x10 Zeiss objective, 0.25 NA. The image of sample is shown in Figure 6.1.

The image in Figure 6.1 was then filtered by using a 25x25 average filter. This makes the original sample image becomes a low resolution image. This image is used to test where the system works for low resolutions images. The filtered image of the sample, low resolution sample image, is shown in Figure 6.2.

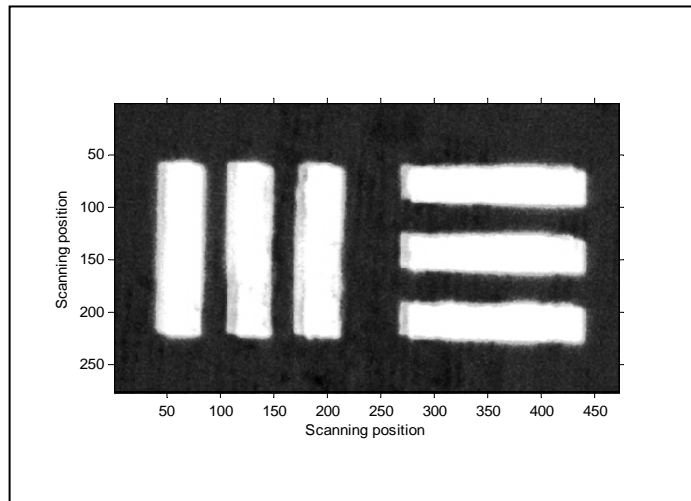


Figure 6.1 Original sample image

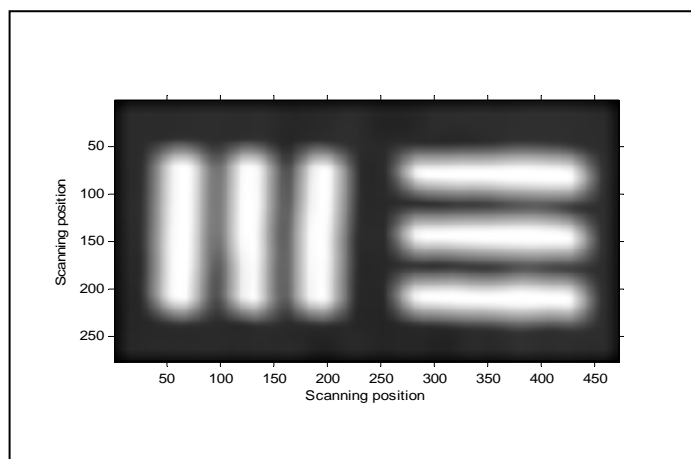
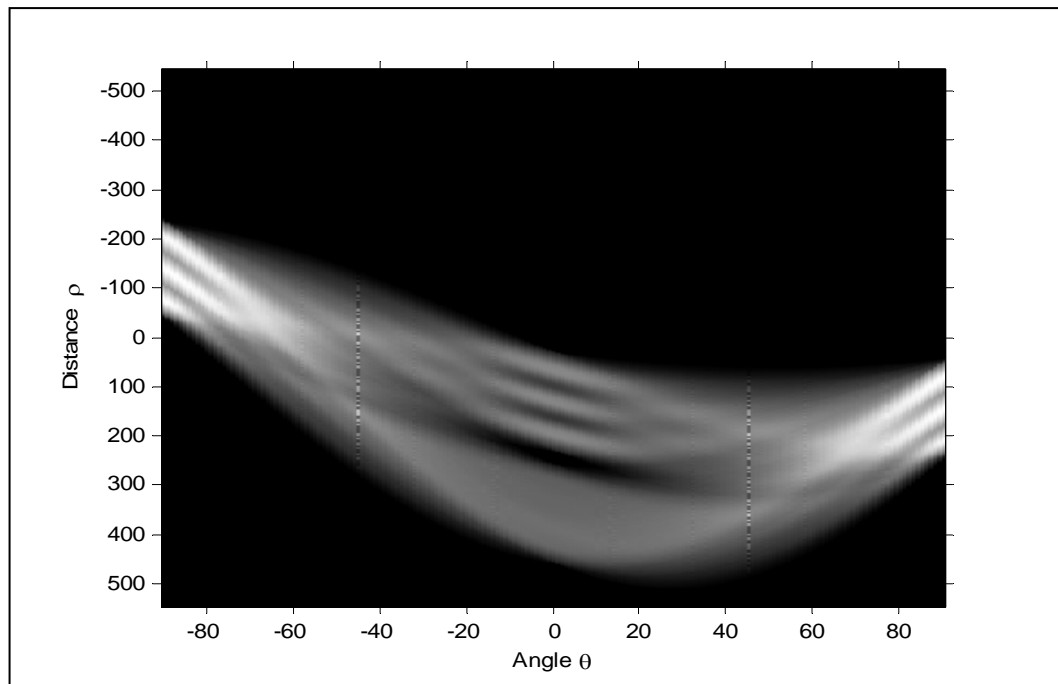
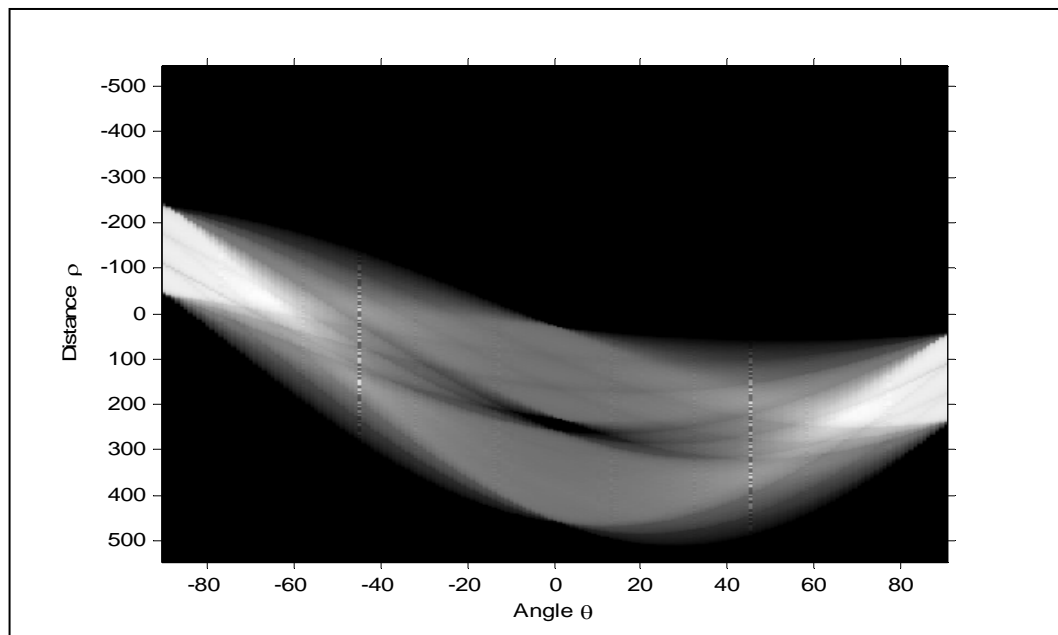


Figure 6.2 Low resolution sample image, original image from Figure 6.1

6.2 Hough Transform

After the sample image (filtered image in Figure 6.2) was obtained, the HT was performed. The following is a step-by-step example of process to obtain parameters from performing HT. The results from performing HT are shown in Figure 6.3 for *Acc1* and Figure 6.4 for *Acc2*.

Figure 6.3 Hough Transform space of *Acc1*Figure 6.4 Hough Transform space of *Acc2*

Determine Peak Areas

The peak areas are obtained by searching the Hough space *Acc1* in Figure 6.3.

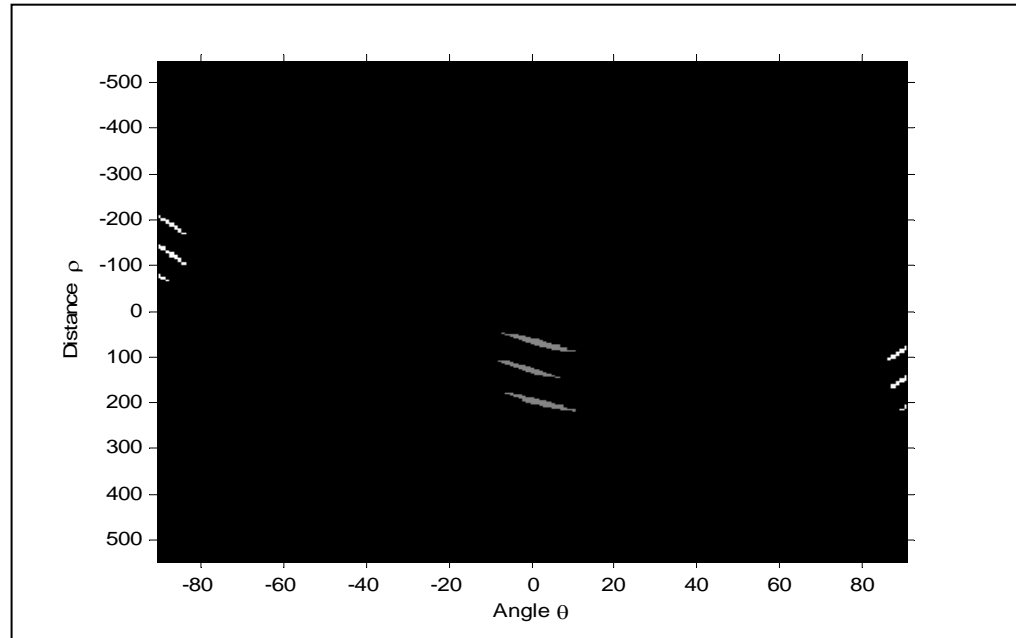


Figure 6.5 Peak areas of *Acc1*

From Figure 6.5, we derive parameters as following:

1. Angle; $\theta = 0^\circ$

Distances ρ for each line from top to bottom are 64, 129 and 194 pixels.

2. Angle; $\theta = 90^\circ$

Distances ρ for each line from top to bottom are 80, 144 and 207 pixels.

It should be noted that the exact orientations of the two groups of line segments are not 0° and 90° . However, for ease of explanation, we will use these two values as indication from now on. From these parameters a profile of line for each angle θ can be drawn. For example, the profile at angle $\theta = 0^\circ$ of *Acc1* is shown in Figure 6.6. It should be noted that this profile is considered only in surrounding area for the

distances ρ of interest, see inside red box in Figure 6.6. The profile at $\theta = 90^\circ$ can be considered in the same way.

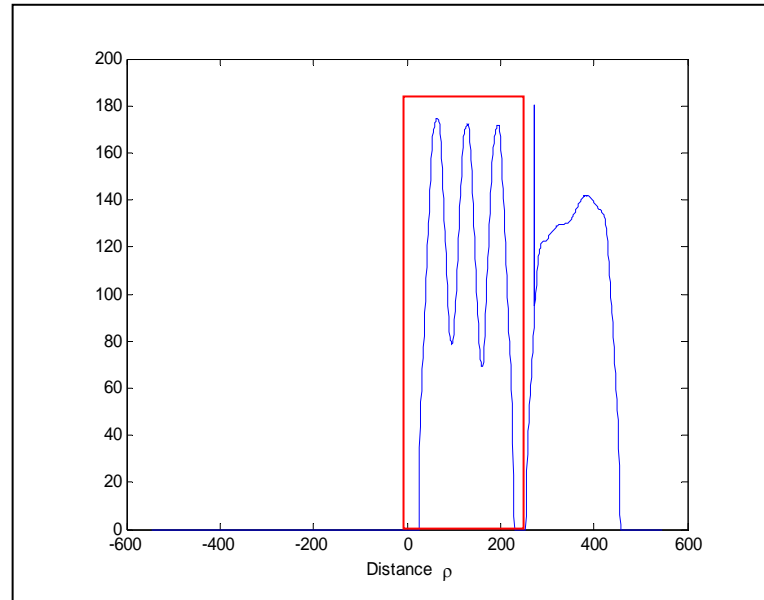


Figure 6.6 Profile taken from *Acc1* at $\theta = 0^\circ$

Determine number of line segments

Number of lines are derived by differentiating the profile in Figure 6.6 and then consider the changes of sign, from positive to negative, in differentiated profile. The differentiated profile is shown in Figure 6.7. Consider Figure 6.7 it can be seen there are three times changes this means there are three line segments in the image at $\theta = 0^\circ$.

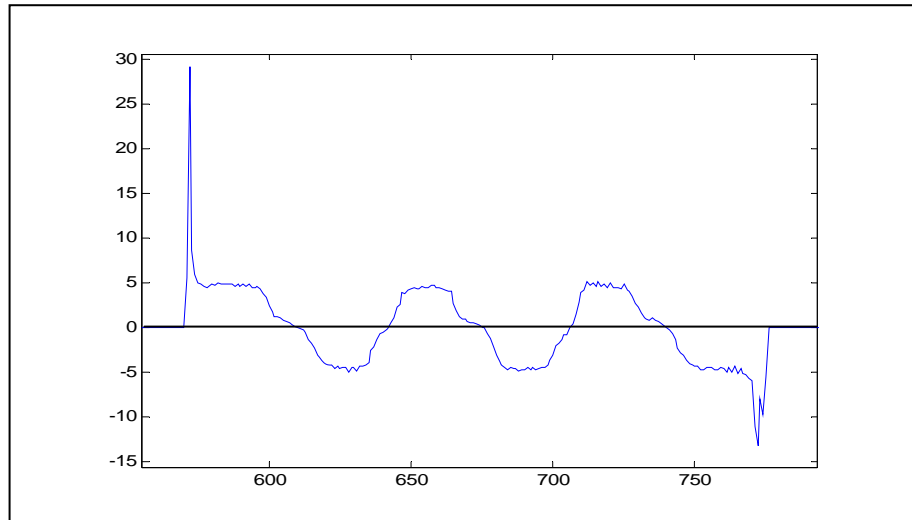


Figure 6.7 Differentiated profile from Figure 6 in red box

The Width and Length of Line

The width and length of a line segment are determined by using *Acc2*. Figure 6.8 shows a profile of *Acc2* (Figure 6.4) taken at the angle $\theta = 0^\circ$. From the section entitled “Determine Peak Areas” above, it is clear that the part of profile that is of interested is within the range $\rho = 0$ to 250. The overall width **W** of the lines is approximately 199 pixels. The lengths of line segments, L1 to L3, are given by the values of the peaks, which are 197, 196 and 196 pixels respectively. The same procedure can be applied to the horizontal line segments in figure 6.2 if desired.

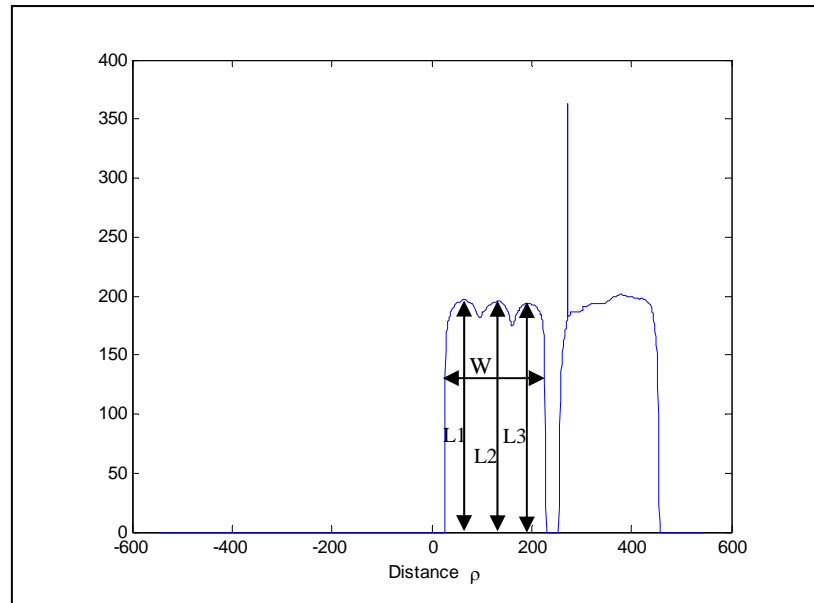


Figure 6.8 Profile taken from *Acc2* at $\theta = 0^\circ$

The parameters from HT are summarised in table 6.1. The units of ρ , width and length are in pixel.

Table 6.1 Parameters of image

| Angle $\theta = 0^\circ$ | | | | Angle $\theta = 90^\circ$ | | | |
|-----------------------------|----------|----------|-------|-----------------------------|----------|----------|-------|
| ρ_1 | ρ_2 | ρ_3 | Width | ρ_1 | ρ_2 | ρ_3 | Width |
| 64 | 129 | 194 | 199 | 80 | 144 | 207 | 200 |
| Length | | | | Length | | | |
| L1 | L2 | L3 | | L1 | L2 | L3 | |
| 197 | 196 | 196 | | 197 | 198 | 197 | |
| Number of line segments = 3 | | | | Number of line segments = 3 | | | |

By performing HT, the approximate parameters of the features of interested are obtained. The width of the individual line segments, however, cannot be determined. The widths W in Table 6.1 are the width for groups of closely spaced line segments, and in the sample used, two groups of such line segments exist. They are the group

of line segments at the angle $\theta = 0^\circ$ and $\theta = 90^\circ$, each group has three line segments. The parameters of each group are known, thus each group can be separated and considered individually. Figure 6.9 shows the separated group of line segments at the angle $\theta = 0^\circ$. This image is then used to be initial image for PCA. Its parameters either from profile or the whole image can be used. The profile is used to measure the width and separation (in case of multiple lines) of the line segments whilst the whole image can be used to measure the length, width and separation.

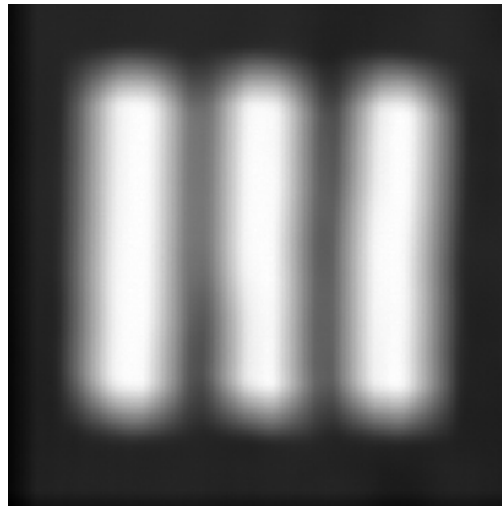


Figure 6.9 Group of line segment at angle $\theta = 0^\circ$

6.3 Principal Component Analysis

After information was obtained from HT, an appropriate template, based on the parameters in Table 6.1 for the line segments at angle $\theta = 0^\circ$, can be selected from the reference data base. Here the template for PCA is a sample of known dimensions. The optical profile of the selected template is shown in Figure 6.10. It should be

noted that the profile is obtained from an image, taken from the same sample used for figure 6.1 but at a different time.

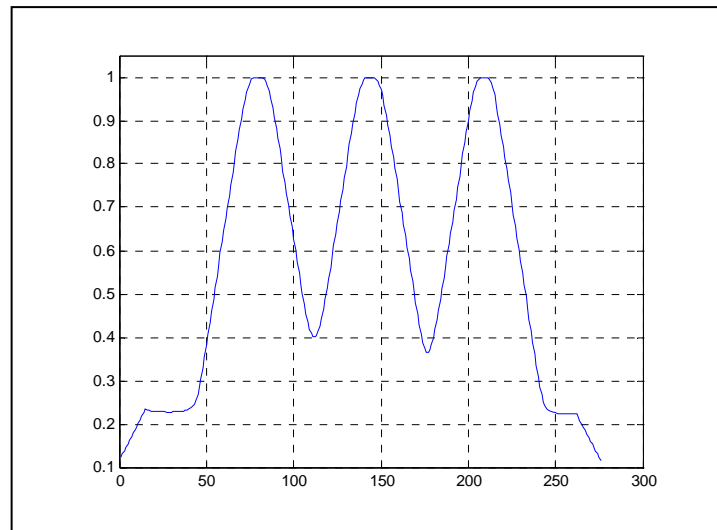


Figure 6.10 Optical profile of the selected template

From the initial image (showing the section at angle $\theta = 0^\circ$) in Figure 6.9, the profile of the sample was taken about the centre of the sample image. The sample profile is shown in Figure 6.11.

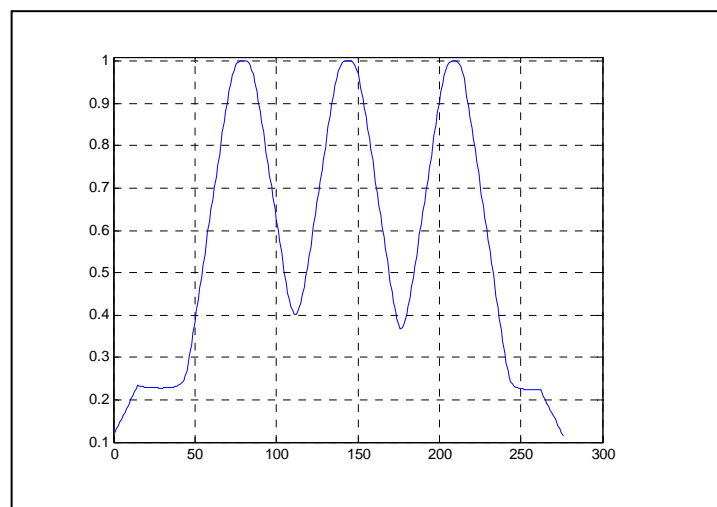


Figure 6.11 Profile of sample

Once profiles of template and sample are obtained, the PCA of these two is performed. The results are tabulated in Table 6.2. The data points plot on PCA axis is shown in Figure 12.

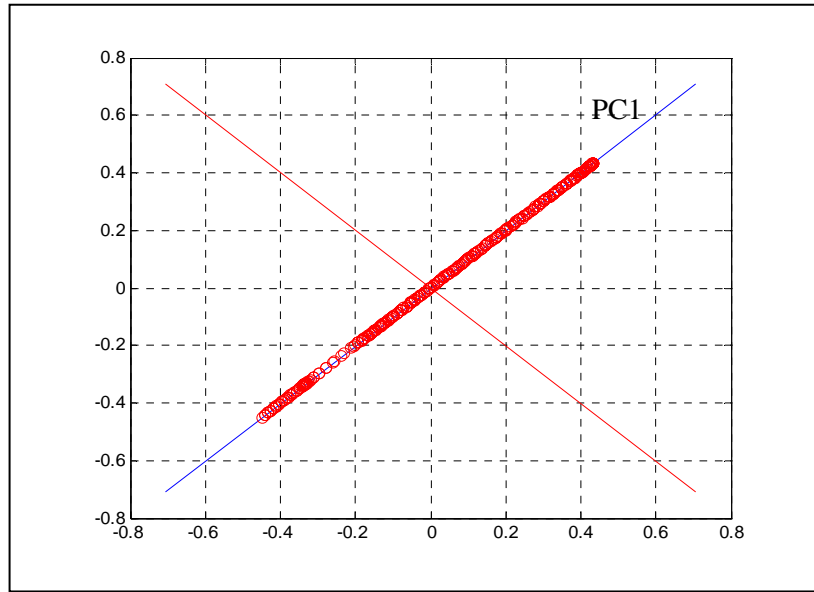


Figure 6.12 Data points on the PCA axis

Table 6.2 PCA result

| V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.7071 | 0.7071 | 0.1737 | 5.068E-7 | 45.00 | 1.00 | 2.91E-6 |

From Table 6.2 it can be seen that the angle $\theta_1 = 45.00^\circ$ this means we obtained the match between the template and sample in all particulars including widths and separations. Note that the ratio of the two eigenvalues is not zero. This is due to the noise associated with the two images. As can be seen in Figure 6.12, all the data points are on the PC1. This means that the template and sample are not shifted.

Performing PCA by using profiles can only measure the width and separation of the line segments in the sample image. Thus in order to measure the lengths of line

segments, the image is used. The sample image is shown in Figure 6.9. A template is selected according to the parameters in Table 6.1, from the reference data base. The sample and template images are then converted to column vectors as described in Chapter 5. These vectors are shown in Figure 6.13 for the sample image and 6.14 for the template image.

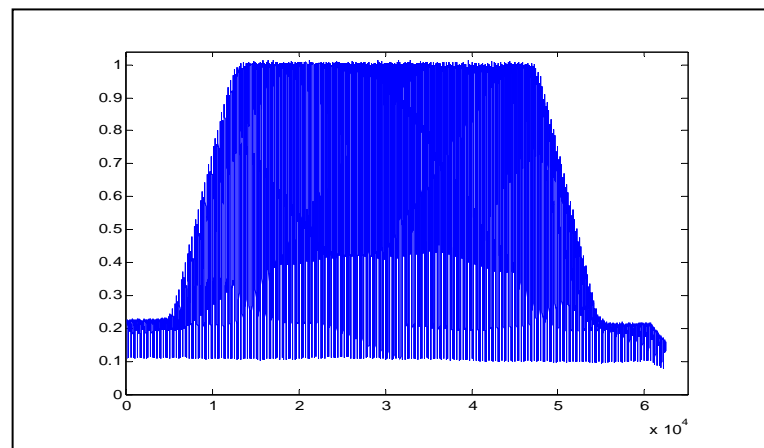


Figure 6.13 The vector of sample image

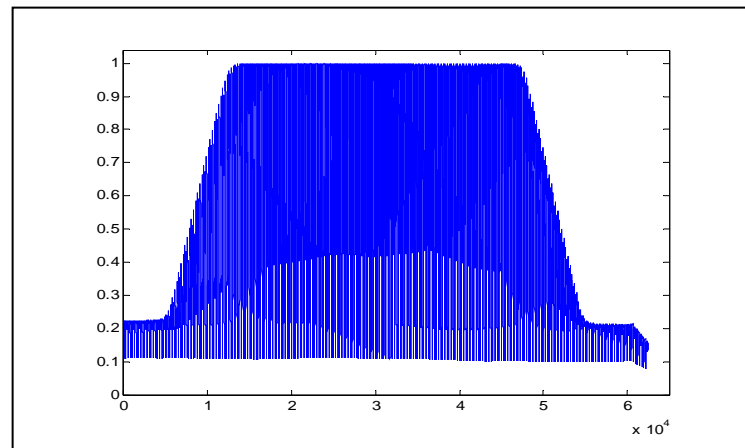


Figure 6.14 The vector of template image

Table 6.3 shows results from performing PCA between the sample and template vectors. From Table 6.3, it can be seen that we have the match with no shifting between the two.

Table 6.3 PCA result

| V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.7071 | 0.7071 | 0.1694 | 5.929E-6 | 45.00 | 1.00 | 3.501E-5 |

The data points plot on PCA axis is shown in Figure 6.15. It can be seen that data points lie on the principal axis. This confirms that the sample and template are matched and there is no shift between sample and template.

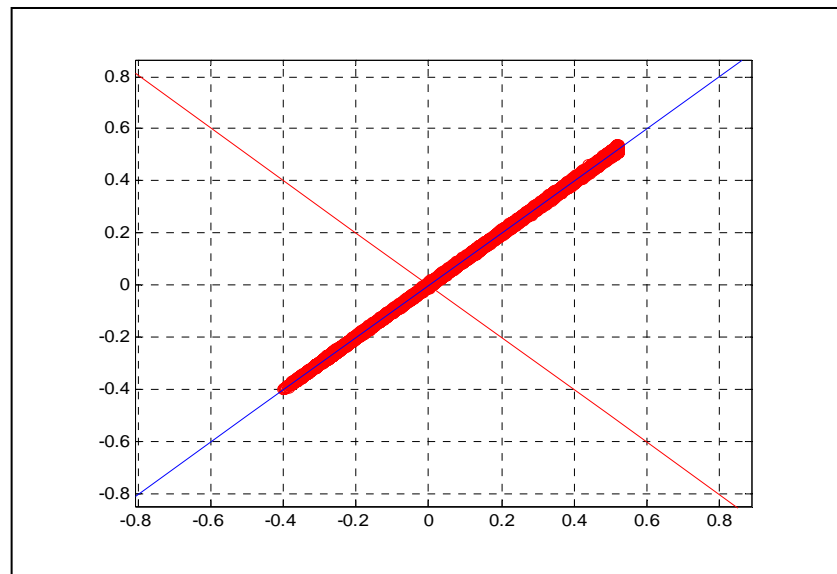


Figure 6.15 Data points on the PCA axis

In summary, the process of identifying and measuring a particular sample has been illustrated in this Chapter. It has been shown that Hough transform can be used to firstly provide an initial identification and measurement of the sample. Based on this information, appropriate templates can be selected, which will serve as the reference for the application of the PCA. The effects of noise on the images can also be seen from the eigenvalues produced from the process.

Chapter 7

7 Conclusion and Future Work

7.1 Conclusion

The main purpose of this research is to reconstruct an object based on a low resolution image of the object with possible applications in semiconductor industry or images of biological samples obtained from special optical techniques, for example, stochastic optical reconstruction microscopy (STORM).

As have been explained in chapter 1, optical systems have limited resolution because of their bandwidths, the nature of diffraction and the wavelength of the illumination sources used. The limited resolution means that objects smaller than approximately 200 nm cannot be resolved optically. Thus being able to extend the resolution range of the optical systems will be very useful for many applications.

Different techniques have been developed to improve the resolution of optical microscopy to beyond the diffraction limit and offer lateral resolution in the nanometers range. These include confocal fluorescence microscopy (CFM), structured illumination microscopy (SIM) and stochastic optical reconstruction microscopy (STORM). In theory, the applications of spectrum extension theory and information theory have been used on the resolution improvement utilising the sampling theorem in the frequency domain and the Auto-regressive model to extend the spectrum of the image beyond the system bandwidth.

Our approach has moved away from the work previously attempted by others where a super-resolved recorded image of the object was the ultimate goal. Our approach is based on the fact that an object image obtained from the optical system has adequate information although the system bandwidth is finite. Therefore the key features of minute object in submicron can be extracted and classified. This approach has become feasible in the recent years due to the rapid advance in the computing power and the availabilities of various signal processing techniques, including principal component analysis, Hough transform and artificial neural networks. We propose to perform classification and extraction of object features. The key parameters or dimensions of these features are then measured, thus allowing object image of much higher resolution to be reconstructed.

Signal Processing

Artificial neural networks

ANN is a powerful signal processing technique which can be trained to perform specific tasks. The strength of ANN in measurement is that it is accurate and robust to noises. However, the ANN demands the large amount of input data and time to train the network. Furthermore, in complex measurement system, many numbers of networks are needed. This means more and more input data and time for training will dramatically increase. Fortunately, by the power of computer the training process becomes effortless. The ANN can be trained easily. Another potential problem with ANN is the selection of the right network for the measurement task required. By the nature of ANN each network is trained so that it is suitable for a narrow range of task

only. Matching the trained ANN with the measurement task is therefore essential for the operation of the system. Both PCA and the Hough transform are used for this purpose.

Hough Transform and Principal component analysis

The Hough Transform and Principal component analysis are used in our project. The system was divided into two stages, coarse and fine selection by using Hough Transform and principal component analysis respectively.

Hough Transform

The HT serves as line segments detector and roughly determines the line image segment parameters. It transforms the image plane into two HT spaces, accumulator *Acc1* and *Acc2*. The difference of the *Acc1* and *Acc2* is that the values of each pixel in the accumulator *Acc1* are the sum of magnitudes of image pixels whereas the *Acc2* are the sum of the number of image pixels of the same distance and angle. By analysing and processing the *Acc1*, numbers of line segments, locations and orientation of individual line image segments are obtained. The *Acc2* is a novel accumulator designed for this project. It is used to measure dimensions, widths and lengths, of the line image segments. Individual lines can be combined to form an image since the parameters of each line image segment is known. This means the shape of the image is known. It must be noted that the HT only provides the initial parameters for the PCA.

Principal component analysis

Based on the initial parameters measured with HT, a proper template is selected for the PCA to perform finer measurement. Various aspects of samples measured by the PCA have been investigated, such as shift of the sample, samples with different sizes, orientation of the sample and the impact of noises.

The main parameters of PCA to be considered when performing PCA between templates and samples are the angle of the eigenvector θ_1 and the ratio of the eigenvalue $\Lambda_{11} / \Lambda_{22}$.

The main result from performing PCA in the case where a template is identical to the sample is presented in Table 7.1.

Table 7.1 Performing PCA result

| V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11}/\Lambda_{22}$ |
|----------|----------|----------------|----------------|------------|-----------------|-----------------------------|
| 0.7071 | 0.7071 | 1977.1 | 0 | 45 | 1 | 0 |

The angle θ_1 of the principal eigenvector is 45° , and the eigenvalue ratio $\Lambda_{11} / \Lambda_{22}$ is zero, corresponding to the eigenvalue Λ_{11} being zero and the other Λ_{22} reaches maximum.

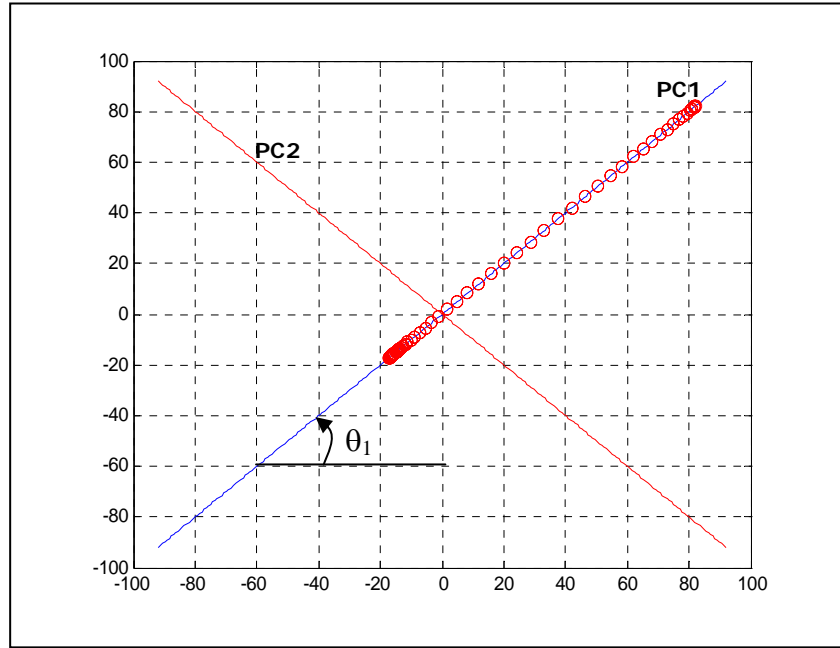


Figure 7.1 A principal component (PC1) and second component (PC2)

Figure 7.1 shows the direction of PC1, and the associated second, orthogonal eigenvector. From the Figure 7.1, it can be seen that all data points lie on PC1.

Performing PCA in cases where the samples are shifted or rotated. The angle θ_1 of the principal vector remains 45° but the eigenvalues ratio $\Lambda_{11}/\Lambda_{22}$ becomes non-zero according to the shift and rotation. This means we can determine the size of template equal to sample, by only checking the angle θ_1 of the eigenvector being 45° . The correct position of the sample can be determine by using the techniques (for each case: shift or rotation) as described in chapter 4 or 5.

In cases where the sizes of the template and sample are different. The PCA can be performed repeatedly, with the size of the template adjusted each time. The targets of

the iteration are the angle θ_1 to approach 45 degrees and also minimising the eigenvalues ratio $\Lambda_{11} / \Lambda_{22}$.

In cases where the system data contains noise. The main result of performing PCA presented in repeated in Table 7.2. From the Table it can be seen that adding noise to the system changes the eigenvalues, with the ratio $\Lambda_{11} / \Lambda_{22}$ decreases as the SNR increases. Noise will also affect the angle θ_1 of the PC1, but only by very small amount. Precisely, the system should have a SNR greater than 80. It should be noted that, for an optical imaging system, a SNR of greater than 250 is readily achieved. Figure 7.2 shows the direction of PC1, and the associated second, orthogonal eigenvector with the SNR=100. From the Figure 7.2, it can be seen that all data points almost perfectly lie on PC1. This is because of the effect of noise.

Table 7.2 Result of PCA

| SNR | V_{12} | V_{22} | Λ_{22} | Λ_{11} | θ_1 | V_{12}/V_{22} | $\Lambda_{11} / \Lambda_{22}$ |
|-----|----------|----------|----------------|----------------|------------|-----------------|-------------------------------|
| 10 | 0.7048 | 0.7093 | 1.89E3 | 8.63 | 45.18 | 0.993 | 4.34E-3 |
| 20 | 0.7063 | 0.7078 | 3.2E4 | 35.10 | 45.06 | 0.998 | 1.10E-3 |
| 30 | 0.7069 | 0.7072 | 1.6E5 | 77.98 | 45.013 | 0.999 | 4.85E-4 |
| 50 | 0.7069 | 0.7073 | 1.24E6 | 210.27 | 45.016 | 0.999 | 1.69E-4 |
| 80 | 0.7072 | 0.7069 | 8.12E6 | 541.39 | 44.988 | 0.999 | 6.66E-5 |
| 100 | 0.7071 | 0.7071 | 1.98E7 | 852.67 | 45.00 | 1.00 | 4.30E-5 |
| 150 | 0.7071 | 0.7071 | 1.00E8 | 1950.9 | 45.00 | 1.00 | 1.94E-5 |
| 200 | 0.7071 | 0.7071 | 3.17E8 | 3470.1 | 45.00 | 1.00 | 1.09E-5 |
| 223 | 0.7071 | 0.7071 | 4.95E8 | 4155.6 | 45.00 | 1.00 | 8.38E-6 |
| 316 | 0.7071 | 0.7071 | 1.98E9 | 8859.6 | 45.00 | 1.00 | 4.46E-6 |

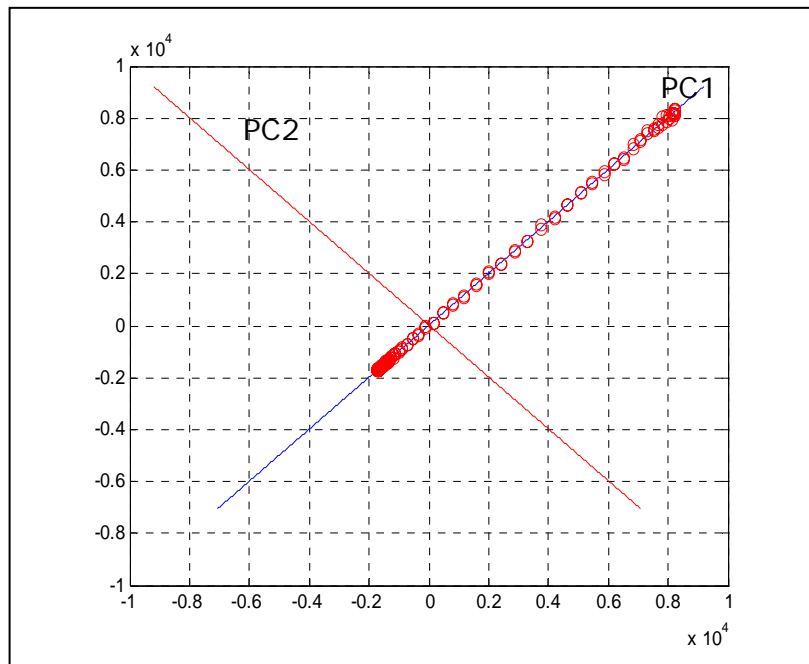


Figure 7.2 Data points on the PCA axis with SNR =100

7.2 Suggestion for Future Work

7.2.1 Multi-Dimension PCA

One interesting future work is to perform multi-dimensional PCA, where a number of templates will be used at any one time. By examining the directions of the eigenvectors and the relative eigenvalues, it is possible to home into the appropriate template more rapidly.

7.2.2 3D Classification

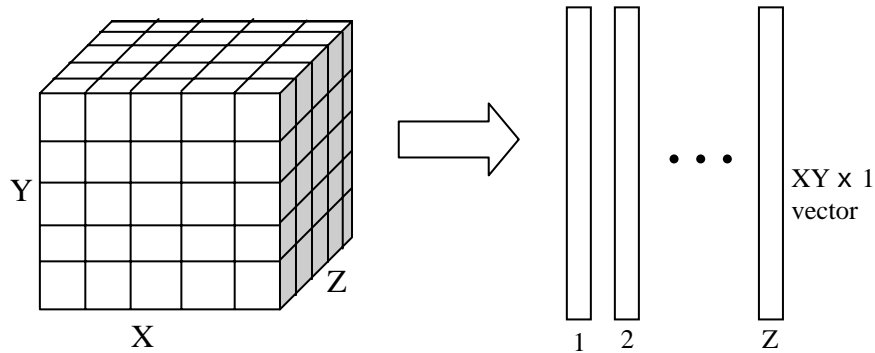


Figure 7.3 Convert 3D image to column vector

With referent to the section 5.1(Convert Line Image to Vector), the 3D image is possible to convert to vector. From the 3D image in Figure 7.3, the image can be converted to many vectors as the length of vector is XY and the number of vector is Z . Each vector of image can be performed PCA one by one as the normal performing PCA. Thus extraction and classification of 3D image features is possible if an optical system provided the 3D image.

7.2.3 Our Vision

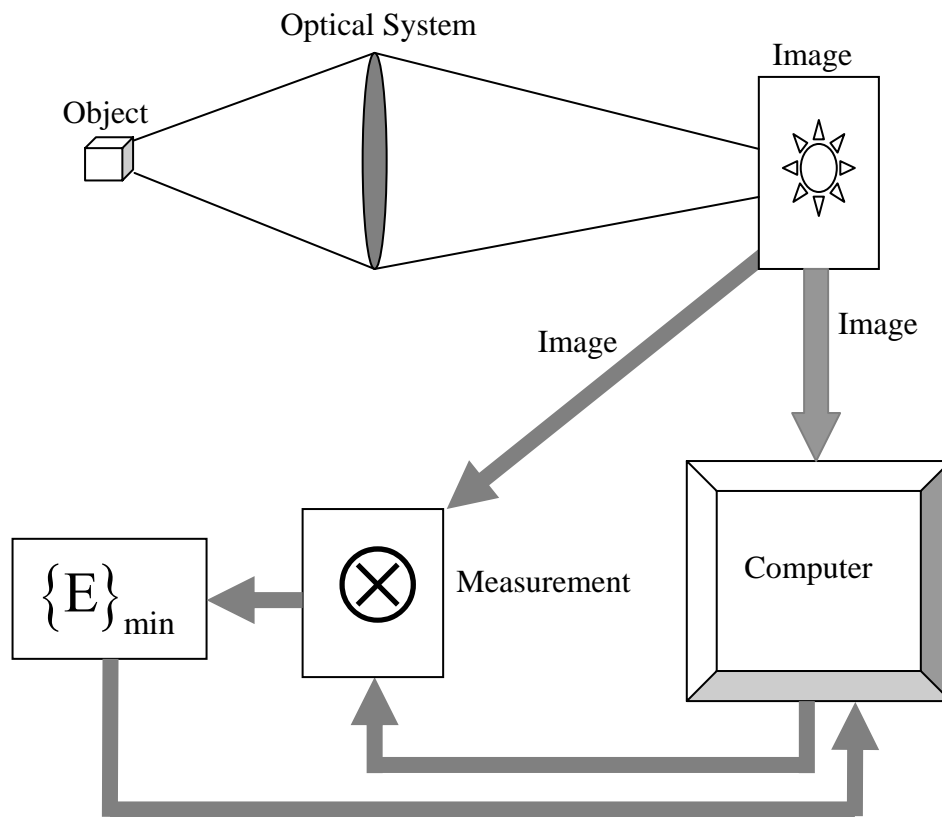


Figure 7.4 The whole computer based model system

By using the power of computer, it is possible to model and reconstruct the whole image of an object from a certain optical system if the features of image of object can be extracted and classified. And then these features are measured to be parameters for the computer system. From Figure 7.4, it can be seen that the computer can be used for extracting and classifying the features of an image. Based on these parameters of the image, a new image can be formed and then compared to the original image. By minimising the error between the reconstructed and original image, the new image can be obtained. This research is only a part of the whole system which tries to extract, classify and measure parameters of low resolution

recorded image from an optical system and then reconstruct the new image with higher resolution.

Appendix A - Eigenvalues Calculation

The eigenvalues can be calculated directly from the distances of data points which are measured perpendicular to PCA axis, PC1 or PC2. The distances d_1 are measured perpendicular to PC1 and d_2 are measured perpendicular to PC2 as shown in Figure A-1.

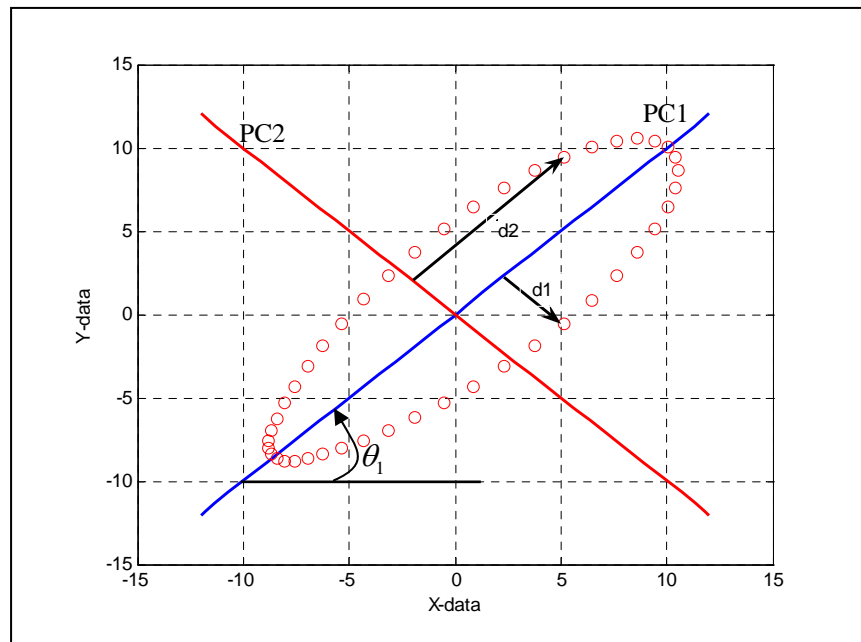


Figure A-1 First Principle Component, PC1, second component, PC2 and data points.

The distances d_1 and d_2 can be calculated from the following equations.

$$d_1 = \left[(X - \bar{x}) - \left(\frac{Y - \bar{y}}{\tan(\theta_1)} \right) \right] \sin(\theta_1)$$

$$d_2 = \left[(X - \bar{x}) - \left(\frac{Y - \bar{y}}{\tan(\theta_2)} \right) \right] \sin(\theta_2)$$

Where \bar{x} and \bar{y} is the mean value of X and Y , respectively. The angle θ_1 is angle of PC1 and angle θ_2 is the angle of PC2. Angles (θ) of vector are simply calculated from eigenvector:

$$\theta_1 = \tan^{-1} \left(\frac{V_{22}}{V_{21}} \right) \quad \text{and}$$

$$\theta_2 = \tan^{-1} \left(\frac{V_{21}}{V_{11}} \right)$$

$$\text{or } \theta_2 = \theta_1 + 90^\circ.$$

The maximum and minimum eigenvalues can be calculated.

Minimum eigenvalue:

$$Eval_{\min} = \frac{1}{n-1} \sum_{i=1}^n (d_{1i} - \bar{d}_1)^2$$

where \bar{d}_1 is mean value of d_1

Maximum eigenvalue:

$$Eval_{\max} = \frac{1}{n-1} \sum_{i=1}^n (d_{2i} - \overline{d_2})^2$$

where $\overline{d_2}$ is mean value of d_2

References

1. Forest preserve district council, *history of microscope*, 20/11/2010
www.newton.dep.anl.gov/nat61tn/500-599/nb506.htm.
2. About, timeline – history of microscopes, 12/02/2011
<http://inventors.about.com/od/mstartinventions/a/microscopes.htm>
3. H Smith, history, 12/02/2011
http://bama.ua.edu/~hsmithso/class/bsc_656/websites/history.html
4. Molecular expressions, science optics and you, “pioneers in optics”,
10/01/2011, www.micro.magnet.fsu.edu/optics/timeline/people/lister.html.
5. E. Hecht, Optics, 3th ed., Addison Wesley Longman, New York, 2002.
Chapter 13, p603.
6. E. Hecht, Optics, 3th ed., Addison Wesley Longman, New York, 2002.
Chapter 10, p463.
7. Mic-UK, the history of the microscope, 24/04/2011
www.microscopy-uk.org.uk/intro/histo.html
8. Abramowitz M, Davidson MW (2003-08-01). “Darkfield Illumination”,
17/03/2011, <http://micro.magnet.fsu.edu/primer/techniques/darkfield.html>.
9. F. Zernike, “Phase Contrast, a New Method for the Microscopic Observation of
Transparent Objects”, Physica 9,689-698,1942.
10. Phase Contrast Microscopy, 25/03/2011
<http://micro.magnet.fsu.edu/primer/techniques/phasecontrast/phaseindex.html>
11. Fracon, Progress in microscopy, Pergamon Press, London, 1961

12. M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *Journal of Microscopy* 198, 82-87 (2000).
13. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution", *Science* 313(5793), 1642-1645 (2006).
14. STORM, 03/11/2010, <http://www.hhmi.org/news/betzig.Html>.
15. PALM, 03/12/2010, <http://dx.doi.org/10.1126/science.1127344>
16. PALM, 03/12/2010, <http://zhuang.harvard.edu/>
17. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)", *Vol.3, No.10*, 793-796 (2006).
18. STORM, 25/01/2011, <http://dx.doi.org/10.1038/nmeth929>
19. B. Bailey, D. L. Farkas, D. L. Taylor, F. Lanni, "Enhancement of axial resolution in fluorescence microscopy by standing-wave excitation", *Nature*, 366, 44–48, 1993.
20. M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution", *PNAS*, 102(37), 13081–13086, 2005.
21. M. A. Hayat, "Principles and techniques of electron microscopy: biological applications", 4th ed, Cambridge University Press, 2000.
22. J. L. Harris, "Diffraction and resolving power", *J. Opt. Soc. Am.* 54 (7) 931-936, 1964.

23. C. W. Barnes, "Object restoration in a diffraction limited imaging system", J. Opt. Soc. Am. 56 (5), 575-578, 1966.
24. R. W. Gerchberg, "Super-resolution through error energy reduction", Opt. Acta 21 (9), 709-720, 1974.
25. S.J. Howard, "Method for continuing Fourier spectra given by the fast Fourier transform", J. Opt. Soc, Am. 71(1), 95-98, 1981.
26. R.J. Smith, C.W. See, M.G. Somekh, A. Yacoot and E. Choi, "Optical track width measurements below 100 nm using artificial neural networks", IOP Meas. Sci. Technol, 16, 2397-2404, 2005.
27. R.J. Smith, C.W. See, M.G. Somekh, and A. Yacoot , "Use artificial neural networks on optical track width measurements", Optical Society of America, Applied Optics, Vol.46, No.22, 2007.
28. H. H. Hopkins, "The airy disc formula for systems of high relative aperture", Proceedings of the Physical Society 55, 0116-0128, 1943.
29. E. Hecht, Optics, 4th ed., Addison Wesley Longman, New York, 2002.
30. R. Shannon J Wyant, Applied Optics and optical engineering, Academic Press, New York, p140, 1983.
31. G Toraldo di Francia, Resolving power and information, J. Opt. Soc. Am., Vol. 45, No. 7, 1955.
32. E A Guillemin, the mathematics of circuit analysis, J Wiley, New York, p288,290, 1951.
33. E Whittaker G Watson, A Course of Modern Analysis, Cambridge university press, Cambridge, 1973.
34. W. Lukosz, Optical system with resolving powers exceeding the classical limit, J. Opt. Soc. Am. 56(11), 1463-1472, 1966.

35. W. Lukosz, Optical Systems with Resolving Powers Exceeding the Classical Limit. II, JOSA, Vol 57(7), 932 – 941, 1967.
36. G. Toraldo di Francia, Degrees of Freedom of an Image, JOSA, Vol 59 (7), 799 – 804, 1969.
37. I Cox and C Sheppard, information capacity and resolution in an optical system, J. Opt. Soc. Am., Vol 3, No 8, pp1152-1158, 1986.
38. C.J.R. Sheppard, and K.G.Larkin, Information Capacity and Resolution in Three-Dimensional Imaging, Optik, Vol 113 (12), 548 – 550, 2003.
39. Minami et al, Superresolution of Fourier transform spectra by autoregressive model fitting with singular value decomposition, Applied Optics, Vol 24, No 2 1985.
40. Gurney, an introduction to neural networks, UCL press limited, London, 1997.
41. Schalkoff, Artificial neural networks, McGraw-Hill,London,1997.
42. Zygo corporation, 17/06/2011, www.zygo.com
43. Olympus,LEXT, 17/02/2011,
http://www.olympus-europa.com/medical/26_LEXT.htm
44. Hough and P. V. C, “Method and Means for Recognizing Complex Patterns”, U.S. Patent No. 3069654, 1962.
45. Duda, R. O. and P. E. Hart, “Use of the Hough Transformation to Detect Lines and Curves in Pictures”, Comm. ACM, Vol. 15, pp. 11–15, 1972.
46. D. H. Ballard, “Generalizing the hough transform to detect arbitrary shapes”. In Readings in Computer Vision: Issues, Problems, Principles, and Paradigms,714-725, 1987.
47. J. Princen, H. K. Yuen, J. Illingworth and J. Kittler, “Properties of the adaptive

- Hough Transform”, Proceedings of the 6th Scandinavian Conference on Image Analysis, Oulu, Finland, 1989.
48. K. Pearson, “On Lines and Planes of Closest Fit to Systems of Points in Space” 24/09/2010, ([http:// stat. smmu. edu. cn/ history/ pearson1901.pdf](http://stat.smmu.edu.cn/history/pearson1901.pdf)) (PDF). Philosophical Magazine 2 (6): 559–572, 1901.
 49. Hotelling, “Analysis of a complex of statistical variables into principal components”, Journal of Educational Psychology, 25, 417-441, 1933.
 50. Jolliffe, I.T, “Principal component analysis”, New York: Springer, 2002.
 51. Jackson, J.E, “A user's guide to principal components”, New York: Wiley, 1991.
 52. Saporta, G, and Niang, N, “Principal component analysis: application to statistical process control”, In G. Govaert (Ed), Data analysis. (pp. 1{23). London: Wiley, 2009.
 53. K. Kim, “ Face recognition using Principal component Analysis”, Department of Computer Science, University of Maryland, College Park, MD 20742, USA.
 54. M. Turk, A. Pentland, “ Eigenfaces for recognition”, Journal of cognitive neuroscience, Vol.2, No. 3, 1991.
 55. W. Qu, X. Huang and Y. Jia, “Segmentation in Noisy Medical Images Using PCA Model Based Particle Filtering”.
 56. A. K. Mishra and B. Mulgrew. “Radar Signal Classification using PCA-based Features”, ICASSP, IEEE, 2006.
 57. Imran S. Bajwa, M. Shahid Naweel , M. Nadim Asif and S. Irfan Hyder, “Feature Based Image Classification by using Principal Component Analysis”, ICGST-GVIP Journal, ISSN 1687-398X, Vol.9, Issue II, 2009.

58. Herve Abdi and Lynne J. Williams, “ Principal Component Analysis”, Wiley Interdisciplinary Reviews: Computation Statistics,2, in press, 2010.
59. An introduction to CCD operation, 09/02/2011
http://www.mssl.ucl.ac.uk/www_detector/opttheory/ccdoperation.html
60. CCD detectors, 25/02/2011, <http://astrosurf.com/re/chip.html>
61. SPM - Scanning Probe Microscopy, 05/09/2010,
<http://www.mobot.org/jwccross/spm/>
62. G. Schitter, M. J. Rost (2008). “Scanning probe microscopy at video-rate” (<http://www.materialstoday.com/view/2194/scanning-probe-microscopy-at-videorate/>) (PDF), Materials Today (UK: Elsevier) 11 (special issue): 40–48. doi:10.1016/S1369-7021(09)70006-9. ISSN 1369-7021.
63. Atomic Force Microscopy resource library,19/01/2011,
<http://www.afmuniversity.org>
64. Soo Chang Pei and Ji-Hwei Horng, “Circular arc detection based on Hough Transform”, Pattern recognition letters 16 (1995), 615-625, Elsevier Science B.V. SSDI 0167-8655(95)00010-0.
65. Richard Smith, “Optical measurement of ultra fine linewidths using artificial neural networks”, PhD thesis, the University of Nottingham, August 2006.
66. M. Colleen Gino, “ Noise, Noise, Noise”, 10/03/2012,
<http://astrophys-assist.com/educate/noise/noise.htm>
67. R Heintzmann and G Ficz, “Breaking the resolution limit in light microscopy”, Briefings in functional genomics and proteomics, vol 5, no 4, 289-301, 2006.
68. Scanning electron microscope, 28/02/2012,
[“http://en.wikipedia.org/wiki/Scanning_electron_microscope”](http://en.wikipedia.org/wiki/Scanning_electron_microscope).

69. Y. Oshikane et al. "Observation of nanostructure by scanning near-field optical microscope with small sphere probe", (http://www.iop.org/EJ/article/1468-6996/8/3/A12/STAM_8_3_A12.pdf) (free pdf).Sci.Technol. Adv.Mater 8:181.doi:10.1016/j.stam.2007.02.013.
70. Neural networks as cybernetic system 2nd and revised edition, Holk Cruse, <http://www.brains-minds-media.org/archive/615/bmm615.pdf>.
71. Salatas, John (24 August 2011), "Implementation of Competitive Learning Networks for WEKA, 10/03/2012, <http://jsalatas.ictpro.gr/implementation-of-competitive-learning-networks-for-weka/> , ICT Research Blog.
72. Resolving power test target USAF-1951, Edmund Optics Inc.